NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis [ID1501]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis [ID1501]

Contents:

The following documents are made available to consultees and commentators:

The **final scope** and **final stakeholder list** are available on the NICE website.

1. **Company submission** from **Celgene, a BMS company**
   a. Evidence submission
   b. Addendum to evidence submission

2. **Clarification questions and company responses**
   a. Clarification response – August 2020
   b. Additional Clarification response – August 2020
   c. Additional Clarification response – August 2020
   d. Clarification response for addendum to evidence submission – March 2021

3. **Patient group, professional group and NHS organisation submission** from:
   a. Leukaemia Care
   b. MPN Voice

4. **Evidence Review Group report** prepared by **School of Health and Related Research (ScHARR)**
   a. ERG report
   b. ERG critique of company addendum

5. **Evidence Review Group – factual accuracy check**

6. **Technical engagement response** from **Celgene, a BMS company**
   a. Response form
   b. TE response appendix

7. **Technical engagement responses & expert statements from experts:**
   a. Claire Harrison – clinical expert, nominated by Celgene, a BMS company
   b. Adam Mead – clinical expert, nominated by MPN Voice
   c. Mark Rutherford – patient expert, nominated by MPN Voice
   d. Caroline Thomas – patient expert, nominated by MPN Voice

8. **Technical engagement response from consultees and commentators:**
   a. Leukaemia Care

b. MPN Voice

9. Evidence Review Group critique of company response to technical engagement prepared by **School of Health and Related Research (ScHARR)**

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501

Document B

Company evidence submission

July 2020

File name	Version	Contains confidential information	Date
		Yes	**10th July 2020**

Contents

Tables and figures

B.1. Decision problem, description of the technology and clinical care pathway

B.1.1. Decision problem

The marketing authorisation for fedratinib (INREBIC®) is for [redacted]This submission focuses on part of the technology's marketing authorisation for patients who have been treated with ruxolitinib (JAKAVI®). The proposed position in the treatment pathway is narrower than the marketing authorisation because:

- The position reflects the unmet need within the myelofibrosis treatment pathway and reflects where clinicians anticipate using fedratinib in UK practice due to the current lack of active treatments available
- This position provides the most clinical benefit given the poor outcomes, including survival, currently observed with patients who are relapsed, refractory, or intolerant to ruxolitinib
- This position optimises the cost-effectiveness of fedratinib because it shows clinical efficacy in a patient population who continue suboptimal ruxolitinib treatment despite the associated poor outcomes

The decision problem addressed is summarised in Table 1.

Table 1: The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	Adults with primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis	Adults with primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis that have been treated with ruxolitinib	This position reflects where fedratinib provides the most clinical and cost-effectiveness, given that there are currently no other treatment options in this population
Intervention	Fedratinib 400 mg	Fedratinib 400 mg	Not applicable
Comparator(s)	No previous treatment with ruxolitinib • Ruxolitinib (for people with intermediate-2 risk or high-risk disease) • Established clinical practice (including but not limited to hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin and RBC transfusion) Previous treatment with ruxolitinib or if ruxolitinib is not appropriate • Established clinical practice (including but not limited to hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin, and RBC transfusion)	Previous treatment with ruxolitinib or if ruxolitinib is not appropriate • Established clinical practice, otherwise referred to as BAT (including but not limited to ruxolitinib, hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin, and RBC transfusion)	The established clinical practice for patients treated with ruxolitinib in the UK includes treatment with BAT; a basket of treatment options that are supportive and do not alter the course of disease. BAT options largely align with those specified in the NICE scope, with the addition of ruxolitinib. A lack of treatment options for patients who are relapsed or refractory to ruxolitinib means that patients continue to receive suboptimal treatment with ruxolitinib

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Outcomes	The outcome measures to be considered include: • Spleen size • Symptom relief (including itch, pain and fatigue) • Overall survival • Progression-free survival • Response rate • Haematological parameters (including RBC transfusion and blood count) • Adverse effects of treatment • Health-related quality of life	The outcome measures to be considered include: • Spleen size • Symptom relief (including itch, pain and fatigue) • Overall survival • Response rate • Haematological parameters (including RBC transfusion and blood count) • Adverse effects of treatment • Health-related quality of life	Progression-free survival has not been included as an outcome because there is no standardised definition of progression in myelofibrosis and, therefore, it is not a measure used in any clinical trials
Economic analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year The reference case stipulates that the time horizon for estimating clinical and cost-effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared	The reference case has been adhered to (Section B.3.2).	Not applicable.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
	Costs will be considered from an NHS and Personal Social Services perspective. The availability of any patient access schemes for the intervention or comparator technologies will be taken into account		
Key: BAT, best available therapy; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; RBC, red blood cell transfusion.			

B.1.2. Description of the technology being appraised

A summary description of fedratinib, including details of its mechanism of action and marketing authorisation, is provided in Table 2.

Appendix C provides a draft summary of the product characteristics.

Table 2: Technology being appraised

UK approved name and brand name	Fedratinib (INREBIC®)
Mechanism of action	Fedratinib is an oral kinase inhibitor with activity against wild-type and mutationally activated JAK2. Fedratinib selectively inhibits JAK2, with higher inhibitory activity for JAK2 over family members JAK1, JAK3 and TYK2. Fedratinib is a more selective inhibitor of JAK2 than ruxolitinib which inhibits both subtypes, JAK1 and JAK2. Abnormal activation of JAK2 is associated with myeloproliferative neoplasms, including primary myelofibrosis, essential thrombocythaemia and polycythaemia vera. In cell models expressing mutationally active JAK2, fedratinib reduced phosphorylation of STAT proteins, inhibited cell proliferation and induced apoptotic cell death. In mouse models of JAK2-driven myeloproliferative disease, fedratinib blocked phosphorylation of STAT 3/5 and improved survival, white blood cell counts, haematocrit, splenomegaly and bone marrow fibrosis.
Marketing authorisation	A marketing authorisation application for the indication below was submitted to the EMA in [redacted]. The anticipated date of CHMP positive opinion is [redacted] and the anticipated date of regulatory approval is [redacted].
Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	The anticipated indication for fedratinib is: [redacted] [redacted] 10^9 /L at [redacted]
Method of administration and dosage	Fedratinib is administered orally as a single daily dose of 400 mg (four 100 mg tablets) taken with or without food
Additional tests or investigations	Thiamine levels in patients should be assessed before starting treatment with fedratinib and during treatment as clinically indicated (e.g. each month for the first 3 months and every 3 months thereafter). Fedratinib treatment should not be started in patients with thiamine deficiency

List price and average cost of a course of treatment	£[redacted]
Patient access scheme (if applicable)	[redacted]
Key: CHMP, Committee for Medicinal Products for Human Use; EMA, European Medicines Agency; JAK, Janus kinase; STAT, signal transducer and activator of transcription.	

B.1.3. Health condition and position of the technology in the treatment pathway

B.1.3.1 Overview of the disease

Myelofibrosis is a rare haematological disorder characterised by abnormal cytopenias, bone marrow fibrosis and extramedullary haematopoiesis; often resulting in splenomegaly, constitutional symptoms and shortened survival.[1, 2] Most patients with myelofibrosis have a mutation that results in constitutive activation of the JAK/Signal Transducer and Activator of Transcription (STAT) signalling pathway.[3, 4] Activation of this pathway results in cell proliferation, inhibition of cell death, and clonal expansion of myeloproliferative malignant cells. The abnormal proliferation of pluripotent haematopoietic stem cells that release inflammatory cytokines and growth factors in the bone marrow leads to marrow fibrosis. Progressive bone marrow fibrosis results in release of the malignant stem cells into the circulation and may result in extramedullary haematopoiesis – manifesting as splenomegaly. Extramedullary haematopoiesis is not able to fully compensate for the loss of production of blood cells in the bone marrow; as a result, patients experience a decrease in one or more blood cell types, i.e. cytopenias (most commonly anaemia and thrombocytopenia). Myelofibrosis may also undergo transformation to acute myeloid leukaemia (AML).[5]

The disease can present as primary myelofibrosis or secondary to polycythaemia vera or essential thrombocythaemia. Myelofibrosis is diagnosed and stratified by risk using one of the following scoring systems – the International Prognostic Scoring System (IPSS), the Dynamic International Prognostic Scoring System (DIPSS) or DIPSS Plus.[6] These are used to classify patients into one of four risk groups (low, intermediate-1, intermediate-2, and high-risk) based on factors such as age, presence of constitutional symptoms, and haematological parameters.

Approximately half of patients with myelofibrosis are found to have either intermediate-2 or high-risk myelofibrosis[7], which is associated with a poor overall prognosis and very limited survival time (see below).[8]

B.1.3.2 Epidemiology and prognosis

Myelofibrosis typically occurs more frequently with increasing age, with the median age at diagnosis being approximately 65 years.[9-11] It impacts slightly more men than women (62%).[9] Epidemiological estimates for myelofibrosis in UK patients suggest a prevalence of 2.2/100,000 and an incidence of 0.4/100,000.[7] This suggests that the total population size of people with myelofibrosis is 1,537, half of which are expected to be intermediate-2 and high risk.

Patients within these risk groups represent a population with considerably worse outcomes compared to patients with intermediate-1 or low risk disease.[5, 13, 14] Currently, only ruxolitinib is recommended by NICE for use in patients with intermediate-2 or high-risk disease.[7] When patients become relapsed, refractory or intolerant to treatment, survival outcomes are poor with several published reports demonstrating a median overall survival (OS) of 13–16 months post ruxolitinib treatment (see Table 47).[7, 15-17] The poor survival outcomes in these patients are attributable to the lack of effective treatment options in the relapsed, refractory and intolerant to ruxolitinib setting, with many patients on suboptimal treatment (see Clinical pathway of care).[15, 16] Data from clinical trials indicate that the majority of patients who are relapsed and refractory to ruxolitinib continue suboptimal ruxolitinib treatment with limited benefits in the absence of other active treatment options.[41,42]

The size of the intermediate-2 and high-risk population who are relapsed, refractory and intolerant to ruxolitinib in the UK is uncertain. The Haematological Malignancy Research Network (HMRN) measured treatment outcomes in [redacted] patients newly diagnosed with primary myelofibrosis ([redacted]%) and secondary myelofibrosis ([redacted]%) between 1 September 2004 and 31 August 2017 in the Yorkshire and the Humber & Yorkshire Coast Cancer Networks.[12] As of 2020, estimates captured in the HMRN analysis indicate that [redacted] patients had initiated ruxolitinib treatment since its EMA marketing authorisation in 2012.[12] It is known the proportion of myelofibrosis patients initiated on, or maintained on ruxolitinib therapy in the UK has changed considerably

since 2017, following the review of ruxolitinib from the Cancer Drugs Fund by NICE.[7] The HMRN figure may not be considered representative of UK clinical practice, given that the estimated uptake of ruxolitinib following its approval and licencing was not observed in HMRN data (see Appendix N). Additionally, there is increasing evidence to suggest that current clinical practice is to maintain patients on ruxolitinib treatment after loss of response (see Section B.1.3.4).

In the HMRN dataset, [redacted] patients discontinued treatment with ruxolitinib, The median time to ruxolitinib discontinuation in these patients was [redacted] years, and the median overall survival from the end of ruxolitinib treatment was [redacted] (Figure 1).

During an advisory board held on 8 April 2020, clinicians substantiated that patients who continue suboptimal ruxolitinib have poor outcomes, including survival. This highlights the need for a new treatment that improves outcomes in patients who are relapsed, refractory and intolerant to ruxolitinib.

Figure 1: Kaplan–Meier estimates of OS from time of discontinuation of ruxolitinib in HMRN patients

Key: HMRN, Haematological Malignancy Research Network; OS, overall survival.

B.1.3.3 Physical and psychological burden of disease

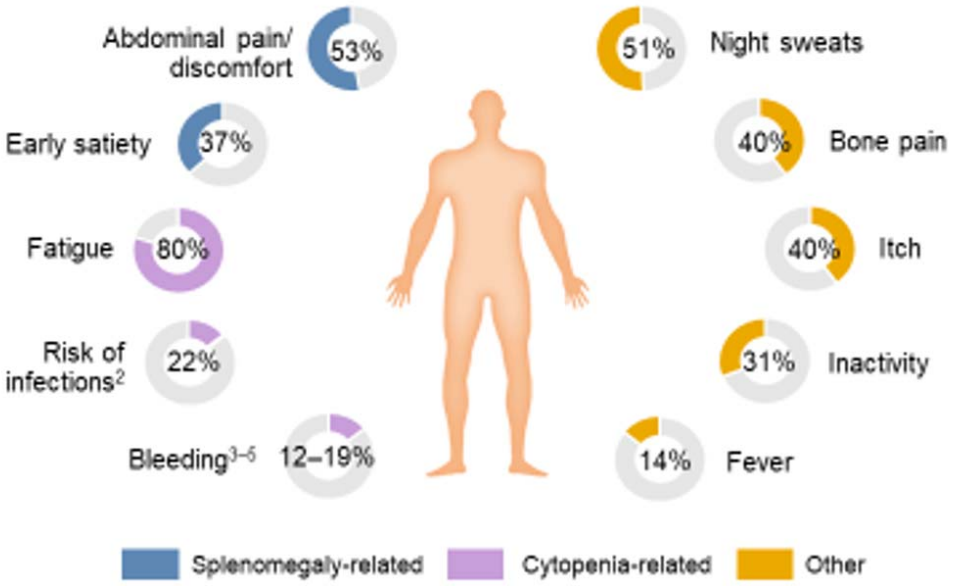
Over 80% of myelofibrosis patients experience splenomegaly, while other clinical manifestations of myelofibrosis include symptoms associated with cytopenias (> 35% of patients), fatigue (> 90%), and constitutional symptoms (~ 30%).[13] Myelofibrosis is associated with a range of debilitating symptoms that may worsen as the disease progresses and can have a major impact on health-related quality of life (HRQoL).[1, 2, 19] These stem from the pathological changes in haematopoiesis and the bone marrow, as described above. Splenomegaly can lead to abdominal pain, early satiety and portal hypertension; while progressive bone marrow fibrosis leads to worsening cytopenias, particularly thrombocytopenia and anaemia.[1] Anaemia is associated with fatigue, weakness, palpitations, bone pain and dyspnoea[1], while cytopenias such as thrombocytopenia and neutropenia result in complications such as petechiae and infection respectively. The risk of cytopenias increases with disease progression, resulting in more severe symptoms and an increased risk of leukaemic transformation.

While extramedullary haematopoiesis predominantly occurs in the spleen and liver, it can also occur in other organs resulting in further complications such as chronic headache, spinal cord compression and pleural effusions.[1]

There are also a range of constitutional symptoms that result from abnormal cytokine production related to the proliferation of progenitor cells. These include fatigue, pruritis, night sweats, fever and cachexia (leading to weight loss) (Figure 2).[1, 2, 20]

Approximately 10–20% of primary myelofibrosis patients will progress to AML.[21] These patients have dismal outcomes, with OS ranging from 3 to 8 months and a 1-year survival rate of 5–10%.[10]

Figure 2: Myelofibrosis is associated with a range of debilitating symptoms



Source: Adapted from Mesa, 2016[20, 22-25]

Studies reporting on the impact of myelofibrosis symptoms on HRQoL suggest that myelofibrosis particularly impacts physical and social function, and this impact increases with disease progression.[19, 20, 26] The negative effect on HRQoL experienced by patients with myelofibrosis is comparable with that reported for patients with recurrent cancer and represents a clinically meaningful reduction in the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30) global health score (GHS) compared with the general population.[19] Many patients reduce their working hours or take early retirement because of myelofibrosis.[20, 27, 28]

In patients that have been treated with ruxolitinib, the physical and psychological burden of myelofibrosis is particularly pronounced. A comparison of the HRQoL at baseline for JAK-naive patients from one of the ruxolitinib pivotal trials, COMFORT-II[29], with baseline data for ruxolitinib-exposed patients included in the fedratinib JAKARTA-2 trial suggests that HRQoL is worse in patients who have been treated with ruxolitinib. Both studies assessed HRQoL using the EORTC QLQ-C30. The GHS score at baseline was 56 for patients naive to ruxolitinib, versus 45 in patients previously exposed to ruxolitinib in JAKARTA-2 (see Section B.2.6). Both were lower than has been reported for the general population, a GHS score of 66.[30]

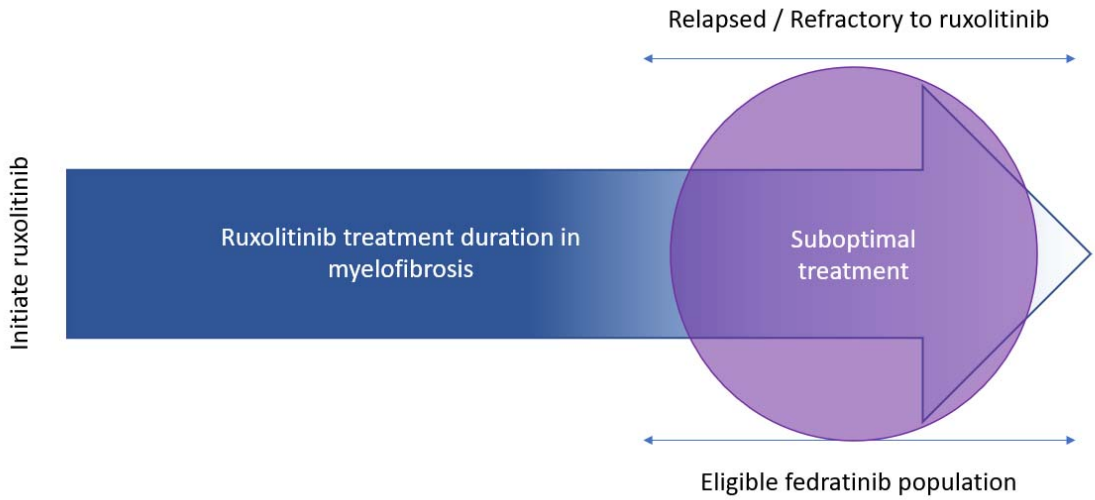
B.1.3.4 Clinical pathway of care

Allogeneic stem cell transplant (ASCT) is the only potentially curative treatment for myelofibrosis; however, it is only suitable for people who are fit enough to undergo treatment as it is associated with considerable morbidity and mortality.[31] ASCT is generally only considered for patients with intermediate-2 or high-risk myelofibrosis, of which only 5–10% will meet eligibility criteria for such an intensive therapy.[32, 33]

Other treatment options aim to relieve debilitating symptoms, particularly splenomegaly and cytopenia, and improve HRQoL. This includes targeted therapy with JAK inhibitors such as ruxolitinib. Ruxolitinib is the only targeted treatment currently approved for myelofibrosis by the European Medicines Agency (EMA) and is used to improve disease-related splenomegaly or symptoms and prolong survival in patients ineligible for curative treatment with ASCT.[34, 35] Similarly, ruxolitinib is the only targeted treatment recommended for use in myelofibrosis patients (with intermediate-2 and high-risk disease) in clinical practice in the UK.[7]

There are considerable limitations associated with treatment with ruxolitinib. Of patients treated with ruxolitinib in clinical trials so far, only 28–42% have achieved the primary endpoint of 35% or more spleen volume reduction (SVR) from baseline.[26, 36, 37] Reports from the COMFORT long-term follow up trials state more than 50% of patients discontinue ruxolitinib treatment after 3–5 years[38], however, this may not be reflective of UK clinical practice. Feedback from UK clinicians at an advisory board revealed that many patients continue to receive suboptimal treatment with ruxolitinib, despite being relapsed or refractory (Figure 3). Reasons for this include the lack of treatment options and concerns regarding the potential for a pro-inflammatory state and acute deterioration of the patient due to ruxolitinib withdrawal.[18, 39] These withdrawal symptoms include acute relapse of disease symptoms, accelerated splenomegaly, worsening of cytopenias, and occasional haemodynamic decompensation (including a septic shock-like syndrome).[40]

Figure 3: Schematic representation of the current treatment duration in those that respond to ruxolitinib



Key: BAT, best available therapy

This continuation of suboptimal ruxolitinib in UK clinical practice aligns with observations from PERSIST-2 and SIMPLIFY-2, where considerable proportions of patients in the BAT arms were receiving ruxolitinib (45% and 89%, respectively).[41, 42] BAT includes treatment options that are largely supportive and do not significantly alter the course of the disease. These may also include treatments such as hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin and red blood cell transfusion.

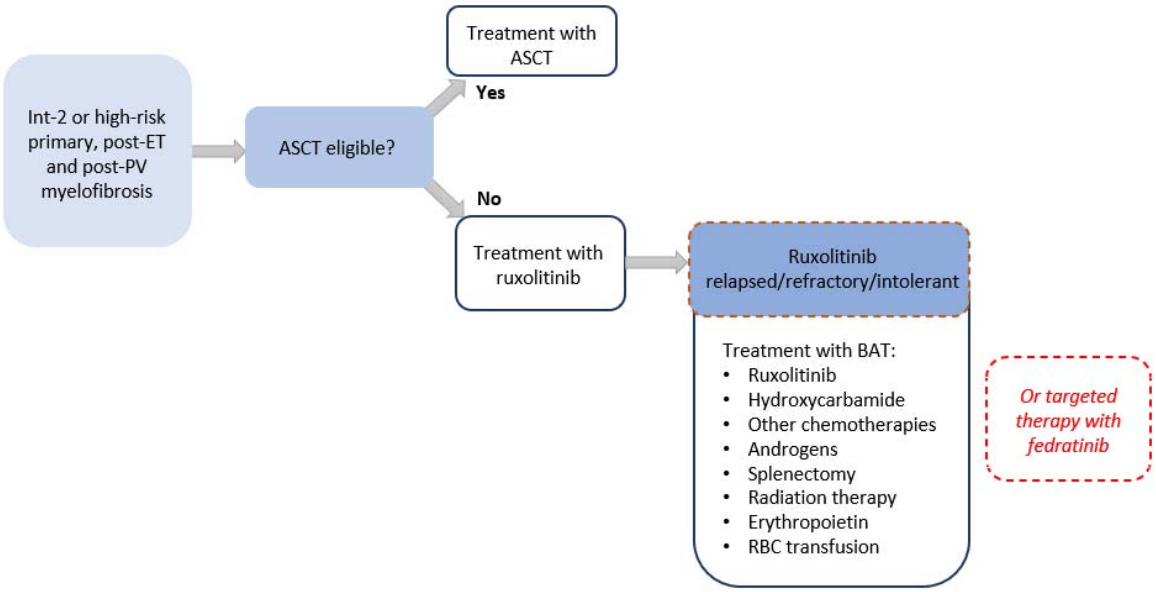
Patients relapsed and refractory to ruxolitinib have a reduced life expectancy, with an estimated median OS of 13–16 months following discontinuation.[7, 15-17] Data on survival in patients who continue suboptimal ruxolitinib is uncertain; however, it is not expected to be significantly greater than observed in the literature, which is supported by clinical experts[18]. There is limited data reported that indicate 21% of patients died at week 24 in those who continue suboptimal ruxolitinib as part of BAT[138].

Given that there are currently not any disease-modifying treatment options available to UK patients no longer responding to ruxolitinib, the introduction of fedratinib to the

pathway of care would provide an opportunity for targeted therapy in a patient population otherwise associated with poor survival outcomes.

The clinical pathway of care for patients with myelofibrosis in England, and potential position of fedratinib within this pathway, is summarised in Figure 4.

Figure 4: Clinical pathway of care for intermediate-2 and high-risk myelofibrosis patients in England



Key: ASCT, allogenic stem cell transplant; BAT, best available therapy; ET, essential thrombocythaemia; Int, intermediate; PV, polycythaemia vera; RBC, red blood cell transfusion.

B.1.3.5 Unmet medical need

In the current clinical pathway of care, ruxolitinib is the only targeted treatment available and is associated with low response rates, with less than half of patients in clinical trials achieving the primary endpoint.[26, 36] In patients that do respond, many will become relapsed or refractory to ruxolitinib over time. In lieu of alternative treatment options, relapsed and refractory patients remain on suboptimal therapy.[18, 39] Outcomes in patients no longer responding to ruxolitinib are poor, with a loss of response associated with worse symptoms and an increased spleen size – causing detriments to HRQoL. There is a significant unmet need for a new therapy to address this and provide an alternative treatment option so that clinicians do not have to resort to using limited healthcare resources for suboptimal treatment.

Fedratinib, a targeted and novel therapy, offers an effective treatment option that has shown a clinically meaningful response in patients who have been treated with ruxolitinib. These benefits lead to considerable HRQoL and survival improvements, in a patient population that would otherwise experience poor outcomes.

B.1.4. Equality considerations

No potential equality considerations have been raised for the use of fedratinib in myelofibrosis patients.

B.2. Clinical effectiveness

B.2.1. Identification and selection of relevant studies

See Appendix D for full details of the process and methods used to identify and select the clinical evidence relevant to the technology being appraised.

In summary, a systematic literature review (SLR) was conducted for primary intervention trials (randomised controlled trials [RCTs] and prospective non-RCTs) assessing the efficacy and safety of fedratinib or comparator therapies in patients with myelofibrosis.

The SLR identified two key studies that evaluated fedratinib as an active intervention:

- The Phase III trial, JAKARTA, investigated the safety and efficacy of fedratinib in the ruxolitinib-naïve population
- The Phase II trial, JAKARTA-2, investigated the safety and efficacy of fedratinib in patients previously treated with ruxolitinib

The SLR also identified studies investigating the use of BAT in patients with myelofibrosis treated with ruxolitinib. These findings have informed the indirect treatment comparison (ITC) of fedratinib versus BAT (see Section B.2.9).

B.2.2. List of relevant clinical effectiveness evidence

The clinical development programme for fedratinib includes two key studies. JAKARTA was a Phase III, double-blind, placebo-controlled study of 289 patients with intermediate-2 or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis. JAKARTA-2 was a Phase II, open-label, single-arm study of 97 patients previously treated with ruxolitinib and with intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis (Table 3).

As JAKARTA-2 provides direct evidence for fedratinib in a patient population who have been treated with ruxolitinib, it forms the key source of clinical and economic evidence in this submission and is described in detail in the following sections.

Further details regarding the results for JAKARTA can be found in Appendix D.

Table 3: Clinical effectiveness evidence

Trial number (acronym)	NCT01523171 (JAKARTA-2)					NCT01437787 (JAKARTA)				
Study design	A Phase II, multicentre, open-label, single-arm study					A Phase III, multicentre, randomised, double-blind, placebo-controlled, three-arm study				
Population	97 patients previously treated with ruxolitinib and with a current diagnosis of intermediate-1 with symptoms, intermediate-2, or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis					289 patients with intermediate-2 or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis with splenomegaly				
Intervention(s)	400 mg fedratinib					400 mg and 500 mg fedratinib				
Comparator(s)	None					Placebo				
Indicate if trial supports application for marketing authorisation	Yes	✓	**Indicate if trial used in the economic model**	Yes	✓	Yes		**Indicate if trial used in the economic model**	Yes	
	No			No		No	✓		No	✓
Rationale for use/non-use in the model	Used in the model as the primary source of evidence for fedratinib in ruxolitinib-exposed patients					Used only where necessary in the model to fill data gaps				
Reported outcomes specified in the decision problem	• **Spleen size** • **Symptom relief** • **Overall survival** • **Response rate** • Haematological parameters • **Adverse effects of treatment** • **Health-related quality of life**					• Spleen size • Symptom relief • Overall survival • Response rate • Haematological parameters • Adverse effects of treatment • Health-related quality of life				
All other reported outcomes	• Duration of response					• Duration of response				

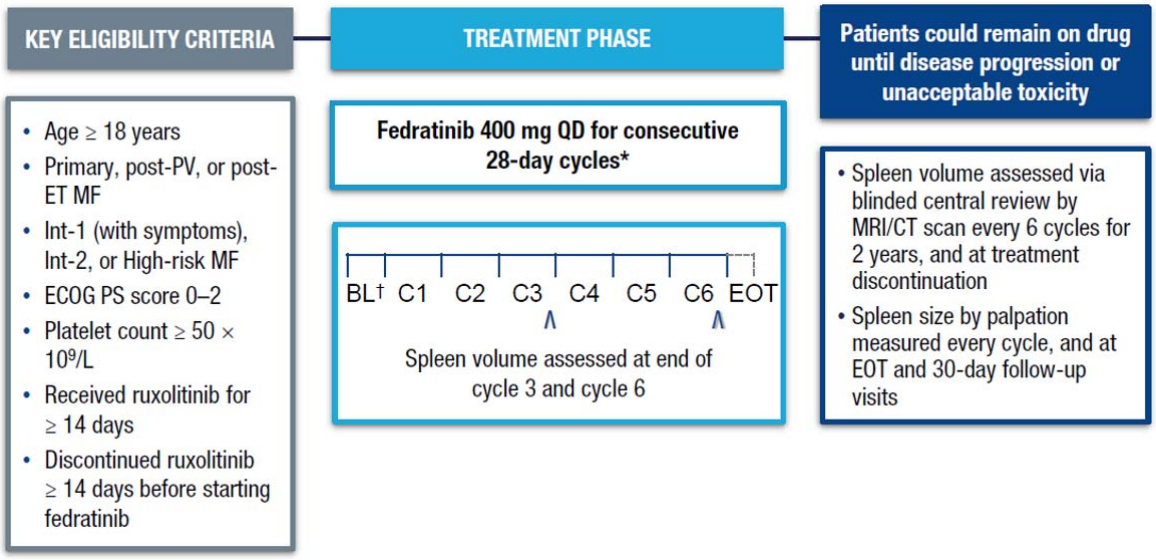
B.2.3. Summary of methodology of the relevant clinical effectiveness evidence

JAKARTA-2 was a Phase II, multicentre, open-label, single-arm study that evaluated the efficacy of a once daily, 400 mg dose of fedratinib in 97 patients previously treated with ruxolitinib.[37] The study included adult patients aged ≥ 18 years with a current diagnosis of intermediate-1 with symptoms, intermediate-2, or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis. Risk categorisation was carried out using the IPSS or DIPSS in patients enrolled after Protocol Amendment 3.

Patients included in JAKARTA-2 were defined as resistant or intolerant to ruxolitinib by investigator assessment.[43] Resistance to ruxolitinib was recorded as either an absence of response, disease progression (increase in spleen size during ruxolitinib treatment) or loss of response at any time during ruxolitinib treatment. Ruxolitinib intolerance was recorded as haematological toxicity (anaemia, thrombocytopenia, other) or non-haematological toxicity. Patients had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib.

The JAKARTA-2 trial design consisted of a screening period of up to 28 days, followed by a treatment phase of six 28-day cycles of fedratinib (24 weeks) and a follow-up visit (approximately 30 days following the last dose of fedratinib).[37] Patients could remain on fedratinib until disease progression or unacceptable toxicity (Figure 5).

Figure 5: JAKARTA-2 study design



Key: BL, baseline; CT, computed tomography; ECOG PS, Eastern Cooperative Oncology Group Performance Status; EOC6, End of Cycle 6; EOT, end of treatment; ET, essential thrombocythaemia; Int, Intermediate; MF, myelofibrosis; MF-SAF, MF Symptom Assessment Form; MRI, magnetic resonance imaging; PV, polycythaemia vera; QD, once daily; TSS, total symptom score.
Notes: *, permitted dose escalation is 400–600 mg/day (dose up-titration permitted if < 50% reduction in spleen size by palpation at the end of Cycles 2 and 4); †, baseline occurred within 14 days of the first fedratinib dose.
Source: Harrison et al. 2019.[44]

The primary outcome measure in JAKARTA-2 was spleen response, defined as the proportion of patients with a ≥ 35% SVR from baseline at the End of Cycle 6 (EOC6).[37] This was measured using magnetic resonance imaging (MRI) or computed tomography (CT) and assessed by blinded central review. Splenomegaly is the main physical feature of myelofibrosis and the cause of many symptoms associated with the disease. As such, SVR is a key treatment goal in myelofibrosis (see Section B.2.13).

Secondary outcomes measured in JAKARTA-2 include[43]:

- Spleen response rate (≥ 35% SVR) at End of Cycle 3 (EOC3)
- Duration of spleen response
- Percent change of spleen volume at EOC3 and EOC6
- Spleen response rate by palpitation at EOC3 and EOC6
- Symptom response rate (≥ 50% reduction in total symptom score [TSS]) at EOC6

Patient-reported outcomes were measured using the Myelofibrosis Symptom Assessment Form (MF-SAF) as an indicator of the effect of fedratinib on symptoms of myelofibrosis and patients' symptom response rates.[43] The EORTC QLQ-C30 was also measured as an exploratory endpoint to capture changes in patients' HRQoL over time. This included measurements of changes to global domains of EORTC QLQ-C30, as well as functional and symptom domains specific to myelofibrosis.

Other clinically relevant exploratory measures included OS and subgroup analyses of the efficacy of fedratinib in patients based on demographic factors and baseline disease characteristics, platelet count at baseline, and patients resistant versus intolerant to ruxolitinib.[43]

The safety of fedratinib was assessed by measuring the incidence of treatment-emergent adverse events (TEAE) and changes from baseline in clinical laboratory parameters and vital signs.[43]

On 14 November 2013, all fedratinib studies (including JAKARTA-2) were put on a clinical hold due to eight suspected cases of Wernicke's encephalopathy (WE) in the fedratinib clinical programme.[45] WE is the presence of neurological symptoms that arise from thiamine deficiency. JAKARTA-2 was subsequently suspended and patients were discontinued from fedratinib treatment and required to initiate thiamine supplementation as a preventative measure.[43] As a result, some patients did not reach EOC6 in the treatment phase of the study.

Based on experts' review, there was a consensus of a clear diagnosis of WE in one out of the eight suspected patients, with the other diagnoses remaining uncertain or inconclusive.[45] WE was found not to be due to a direct pharmacological effect of fedratinib on thiamine absorption or processing, but a consequence of gastrointestinal adverse events (AEs) in undernourished patients.[43] As such, the Food and Drug Administration (FDA) lifted the clinical hold on fedratinib in 2017. The risk of developing WE can be mitigated with routine thiamine monitoring and thiamine replacement, so that patients are able to utilise the clinical benefit offered by fedratinib.

A summary of the methodology of JAKARTA-2 is provided in Table 4.

Table 4: Summary of methodology

Trial number (acronym)	NCT01523171 (JAKARTA-2)
Location	JAKARTA-2 was conducted in 42 sites in nine countries, including one site in the UK
Eligibility criteria for participants	Key inclusion criteria: • Patients who previously received ruxolitinib therapy for the treatment of primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis for at least 14 days (unless the patient discontinued due to intolerance or allergy within 14 days) • Palpable splenomegaly (≥ 5 cm below the left costal margin) • ECOG Performance Status of 2 or less, and life expectancy of 6 months or more Key exclusion criteria: • Received any chemotherapy, including ruxolitinib, within 14 days before the start of the study (except hydroxycarbamide, which was permitted within 1 day of initiation of fedratinib) • A history of other malignancies • Platelet count of < 50 × 10^9 /L
Settings and locations where the data were collected	Steps taken to ensure the accuracy and reliability of the clinical study data included regular site monitoring visits to review study progress, investigator and patient compliance with the protocol requirements, and any emergent problems Data entry and validation were carried out using standard validated remote data capture computer software (Oracle Clinical RDC Version 4.6). Data were stored in an Oracle database on a UNIX server. Data entry was performed directly from the investigator site from the data source documents and signed electronically by the authorised site personnel. Any modification in the database was traced using an audit trail
Trial drugs	400 mg fedratinib was given orally, once daily. If there was a lack of adequate spleen response, the fedratinib dose could be titrated upwards in 100 mg/day increments up to a maximum of 600 mg/day Study treatment continued until disease progression or unacceptable toxicity

Permitted and disallowed concomitant medication	Patients could not receive any other drug treatment for their disease while on study. Treatment with cytotoxic or immunosuppressive therapy, including hydroxycarbamide or systemic corticosteroids (i.e. > 10 mg/day prednisone or equivalent for > 5 days) was prohibited. Use of any other investigational agents during the study was prohibited. The following medications were not to be used prior to inclusion: any chemotherapy, immunomodulatory drug therapy (e.g. thalidomide, interferon-α), anagrelide, immunosuppressive therapy, corticosteroids > 10 mg/day prednisone or equivalent, or growth factor treatment (e.g. erythropoietin), or hormones (e.g. androgens, danazol) within 14 days prior to initiation of fedratinib; and darbepoetin within 28 days prior to initiation of fedratinib
Primary outcome (including scoring methods and timings of assessments)	The primary outcome, spleen response rate, was defined as the proportion of patients with a ≥ 35% SVR at EOC6 relative to baseline, as measured by MRI/CT scan. The MRI/CT scans were reviewed by an independent central imaging laboratory, where reviewers were blinded to the fedratinib doses.
Other outcomes used in the economic model/specified in the scope	Secondary efficacy assessments: • Spleen response rate, defined as the proportion of patients with a ≥ 35% SVR at EOC3, relative to baseline, as measured by MRI/CT scan • Duration of spleen response as measured by MRI/CT • Spleen volume and percent change of spleen volume at EOC3 and EOC6 from baseline as measured by MRI/CT • Proportion of patients with a ≥ 50% reduction in spleen size by palpation at EOC3 and EOC6, relative to baseline • Symptom response rate, defined as the proportion of patients with ≥ 50% reduction in the TSS at EOC6 relative to baseline Key exploratory assessments: • OS, defined as the proportion of patients alive at the time of final analysis • Change in HRQoL using EORTC QLQ-C30 V3.0

Pre-planned subgroups	Analyses of spleen volume reduction and symptom response rate were measured in pre-planned subgroups of: • Demographic factors and baseline disease characteristics • Platelet count at baseline (< 100 x 10^9/L or ≥ 100 x 10^9/L) • Patients resistant versus intolerant to ruxolitinib

Key: CSR, clinical study report; CT, computed tomography; ECOG, Eastern Cooperative Oncology Group; EOC3, end of Cycle 3; EOC6, end of Cycle 6; HRQoL, health-related quality of life; MRI, magnetic resonance imaging; OS, overall survival; SVR, spleen volume reduction; TSS, total symptom score.
Source: Harrison et al. 2017[37] and JAKARTA-2 CSR.[43]

B.2.3.1 Baseline demographics

The demographics and baseline disease characteristics in JAKARTA-2 are representative of a group of patients with advanced myelofibrosis and a high disease burden, with the majority (79.4%) of patients having received ≥ 2 prior anticancer therapies.

Of patients enrolled in JAKARTA-2, there were comparable proportions of men (55%) and women (45%), most patients were White (94.8%)[43] and the median age was 67 years.[37] The largest proportion of patients (55%) had been diagnosed with primary myelofibrosis, followed by post-polycythaemia vera (26%) and post-essential thrombocythaemia (20%).[37] At baseline, the majority of patients had an Eastern Cooperative Oncology Group Performance Status (ECOG PS) of 0 (26.8%) or 1 (46.4%), while 23.7% of patients had an ECOG PS of 2.[43] Almost all patients (95.9%) had constitutional symptoms (night sweats, itching, abdominal discomfort, abdominal pain, early satiety or bone pain) prior to starting treatment with fedratinib.[43] Patients had advanced disease at baseline, with a median baseline spleen volume of 2,894 ml – 12 times that of the normal spleen.[46]

The most frequent myelofibrosis risk categories, as defined by IPSS or DIPSS following a protocol amendment, were intermediate-2 risk (48%) and high-risk (35%), while intermediate-1 risk with symptoms (17%) was less frequent.[37] As ruxolitinib is only recommended by NICE for use in patients with intermediate-2 and high-risk disease, and this submission focuses on patients who have been treated with ruxolitinib, JAKARTA-2 reflects a slightly broader demographic than the target population given that it includes 16 patients with intermediate-1 disease. Analyses of the clinical and cost-effectiveness of fedratinib from JAKARTA-2 have been adjusted to consider removal of these intermediate-1 patients. These analyses demonstrated a clinical benefit that was consistent with the primary analysis.

A summary of the baseline characteristics in JAKARTA-2 is provided in Table 5.

Table 5: Baseline characteristics (JAKARTA-2, ITT population)

	Patients (N=97)
Median age, years (range)	67 (38, 83)
Sex, n (%)	
Male	53 (55%)
Female	44 (45%)
Race, n (%)[a]	
White	92 (94.8%)
Black	1 (1.0%)
Asian	4 (4.1%)
Median weight, kg (range)	73.0 (47.0, 105.7)
Disease type, n (%)	
Primary myelofibrosis	53 (55%)
Post-polycythaemia vera	25 (26%)
Post-essential thrombocythaemia	19 (20%)
Risk status, n (%)[b]	
Intermediate-1	16 (17%)
Intermediate-2	47 (48%)
High-risk	34 (35%)
Median time since diagnosis, years (range)	4.1 (0.3, 24.5)
JAK2 mutational profile, n (%)	
Wild-type	29 (30%)
Mutant	61 (63%)
Missing	7 (7%)
RBC transfusion dependence status, n (%)[c]	
Yes	14 (14%)
No	83 (86%)
Platelet count, n (%)	
<50 x 10^9/L	1 (1%)
≥50 x 10^9/L to <100 x 10^9/L	32 (33%)
≥100 x 10^9/L	64 (66%)
Haemoglobin level, n (%)	
<10 g/dL	51 (53%)
≥10 g/dL	46 (47%)
ECOG, n (%)	
0	26 (26.8%)
1	45 (46.4%)
2	23 (23.7%)
Missing	3 (3.1%)

	Patients (N=97)
Constitutional symptoms[d]	
Yes	93 (95.9%)
No	4 (4.1%)
Median baseline spleen volume, ml (range)	2894 (737, 7815)
Median baseline spleen size, cm (range)[e]	18 (5, 36)

Key: CT, computed tomography; DIPSS, Dynamic International Prognostic Scoring System; ECOG, Eastern Cooperative Oncology Group; IPSS, International Prognostic Scoring System; ITT, intention-to-treat; JAK2, Janus kinase 2; MPN-SAF, myeloproliferative neoplasm symptom assessment form; MRI, magnetic resonance imaging; RBC, red blood cell.
Notes: Spleen volume was measured by MRI/CT scan and reviewed in a blinded fashion by a central imaging laboratory. Spleen size was measured by palpation (i.e. length in cm). [a], the race categories in the electronic case report form were Caucasian/White, Black, Asian/Oriental and other. The race categories in this table were standardised for consistency across fedratinib clinical study reports. Race 'other' is not presented because there were no patients in the category; [b], risk category per IPSS or DIPSS for patients enrolled after Protocol Amendment 3; [c], receiving ≥ 2 units/month of RBC transfusions over 3 months prior to first dose; [d], a subject had constitutional symptoms if any of the symptoms in the baseline MPN-SAF (night sweats, itching, abdominal discomfort, abdominal pain, early satiety, bone pain) had a value greater than zero; [e], below lower coastal region.
Source: Harrison et al. 2017[37], Harrison et al. 2019,[44], Harrison et al. 2020,[46] and JAKARTA-2 CSR.[43]

B.2.3.1.1 Prior myelofibrosis treatment

Patients in JAKARTA-2 were heavily pre-treated, with [redacted]% having received at least two prior anticancer therapies and [redacted]% having received at least four prior anticancer therapies.[43] All 97 patients enrolled in the study had received prior treatment with ruxolitinib, with a median exposure of 10.7 months.[46] Besides ruxolitinib, the most common anticancer therapy was hydroxycarbamide, received by [redacted]% of patients.[43]

Of the patients enrolled and treated in JAKARTA-2, the majority (66%) were resistant to ruxolitinib, a third (33%) were intolerant to ruxolitinib and one patient (1%) was neither resistant nor intolerant and was categorised as 'other: lack of efficacy'.[46]

A summary of the reasons for ruxolitinib discontinuation is provided in Table 6.

Table 6: Reasons for ruxolitinib discontinuation by investigator assessment (JAKARTA-2, ITT population)

	Fedratinib 400 mg (N = 97)
Ruxolitinib resistance, n (%)[a]	64 (66%)
Lack of response	24 (25%)

	Fedratinib 400 mg (N = 97)
Disease progression	15 (16%)
Loss of response	25 (26%)
Ruxolitinib intolerance, n (%)[a]	32 (33%)
Haematological toxicity	25 (26%)
Thrombocytopenia	13 (13%)
Anaemia	9 (9%)
Other	3 (3%)
Non-haematological toxicity	7 (7%)
Other: lack of efficacy, n (%)[b]	1 (1%)

Key: eCRF, electronic case report form; ITT, intent to treat.
Notes: [a], the investigator's opinion was recorded on the relevant eCRF pages as:
– Resistance: lack of response (absence of response), disease progression (spleen size increase during ruxolitinib treatment), loss of response at any time during ruxolitinib treatment;
– Intolerance: haematological toxicity (anaemia, thrombocytopenia, other), non-haematological toxicity. [b], one patient was neither resistant nor intolerant per investigator's assessment and was categorised under 'other: lack of efficacy'.
Source: Harrison et al. 2020[46]

B.2.3.2 JAKARTA-2 reanalysis

JAKARTA-2 was initiated shortly after the approval of ruxolitinib; therefore, the criteria for defining ruxolitinib resistance or intolerance were not yet well defined.[43] Patients in the original protocol were classified as resistant or intolerant to ruxolitinib per the investigators' assessments. A reanalysis of the efficacy of fedratinib in JAKARTA-2 was performed on patients determined to be relapsed or refractory or intolerant to ruxolitinib, based on criteria recommended by myelofibrosis experts from the US and EU at an advisory board meeting and later discussed with health authorities.[47]

These more stringent definitions of ruxolitinib failure are presented in Table 7. The criteria are currently being used in ongoing studies of myelofibrosis in patients that have been treated with ruxolitinib.

Table 7: Ruxolitinib failure criteria

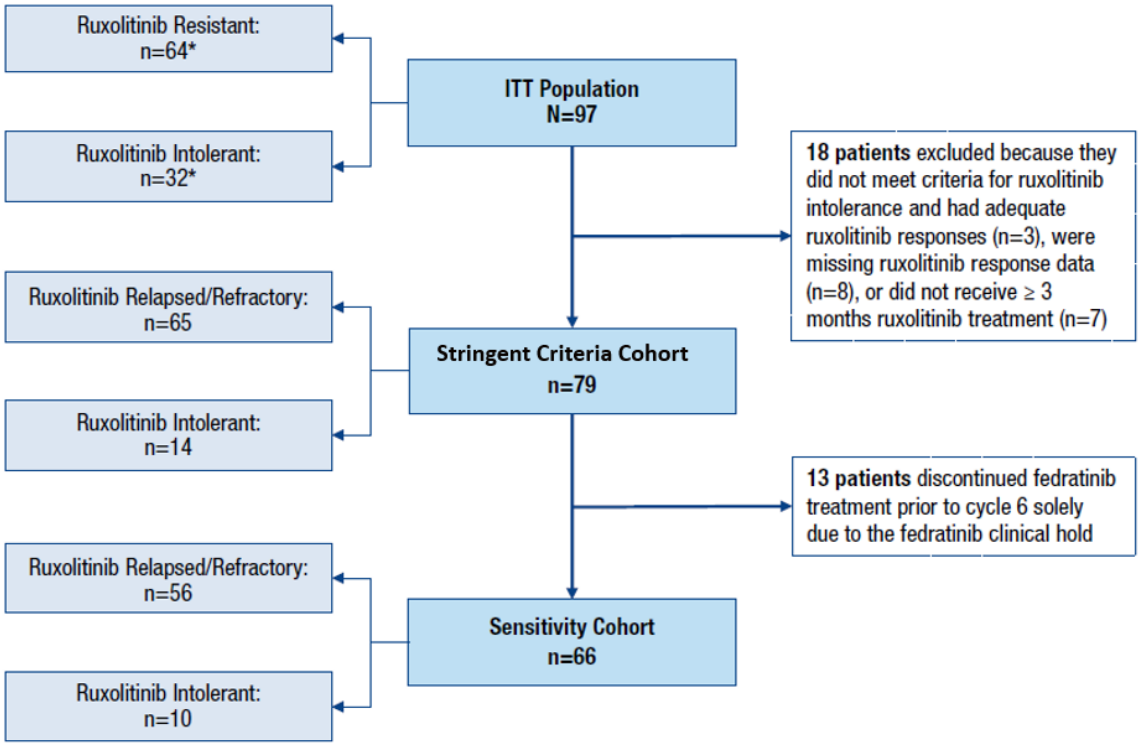
ITT population (N = 97)	Ruxolitinib failure cohort (n = 79)
Ruxolitinib treatment for ≥ 14 days and resistant or intolerant to ruxolitinib per investigator discretion:	Relapsed: ruxolitinib treatment for ≥ 3 months with regrowth, defined as <10% SVR or < 30% decrease in spleen size from baseline, following an initial response

ITT population (N = 97)	Ruxolitinib failure cohort (n = 79)
• Resistant: no response or stable disease, evidence of disease progression or loss of response	Refractory: ruxolitinib treatment for ≥ 3 months with < 10% SVR or <30% decrease in spleen size from baseline
• Intolerant: discontinuation due to unacceptable toxicity	Intolerant: ruxolitinib treatment for ≥28 days complicated by the development of RBC transfusion requirement (≥ 2 units per month for 2 months); or Grade ≥ 3 thrombocytopenia, anaemia, haematoma and/or haemorrhage while receiving ruxolitinib
Key: ITT, intent to treat; RBC, red blood cell; SVR, spleen volume reduction. **Source:** Harrison et al. 2019.[44]	

This analysis split patients into two populations: the Stringent Criteria Cohort comprising 79 patients who met at least one criterion from the stringent definitions for ruxolitinib relapsed, refractory or intolerant; and the sensitivity cohort comprising 66 patients who received six fedratinib treatment cycles or discontinued before EOC6 for reasons other than 'study terminated by sponsor'.[44] The aim of the sensitivity cohort analysis is to estimate fedratinib response without the impact of the clinical hold.

A consort diagram depicting how these criteria were applied to the intent to treat (ITT) population to generate ruxolitinib Stringent Criteria and Sensitivity Cohorts is provided in Figure 6.

Figure 6: Consort diagram (JAKARTA-2, reanalysis)



Key: ITT, intent to treat.
Source: Adapted from Harrison et al. 2019.[44]

The baseline characteristics of the Ruxolitinib Failure and Sensitivity Cohort are provided in Appendix D.

B.2.4. Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

The primary objective of JAKARTA-2 was to determine efficacy of fedratinib with regards to the reduction of spleen volume.[37] Assuming 25% of patients achieved the primary endpoint of ≥ 35% reduction in spleen volume from baseline, 70 evaluable patients were required to provide at least 90% power (at a one-sided 2.5% α level) to test the null hypothesis of ≤ 10% of patients achieving the primary endpoint.

The primary analysis of JAKARTA-2 was conducted in the per protocol (PP) population (n = 83), defined as patients with evaluable baseline and at least one post baseline MRI/CT scan of spleen volume (EOC3 or EOC6) and no important protocol

deviations that could impact the efficacy outcome.[43] In patients who did not reach EOC6 owing to the clinical hold, missing data were accounted for using the last observation carried forward (LOCF) method. As the PP population represents a smaller population than the ITT population, and considering the statistical limitations of the LOCF method, the analyses in the PP population are considered supportive to the ITT population in this submission and are presented in Appendix D.

The ITT population comprised all 97 patients enrolled in the study and provides the largest sample size and statistically robust source for evaluations of efficacy in JAKARTA-2. A reanalysis of JAKARTA-2 data was conducted to confirm the efficacy of fedratinib in subsets of enrolled patients who met new stringent definitions of ruxolitinib relapsed, refractory or intolerant (Figure 6).[46] This reanalysis established that the efficacy of fedratinib is consistent, regardless of the relapse or refractory criteria applied (see Section B.2.6).

In order to determine the treatment effect of fedratinib on clinically important subpopulations, prespecified subgroup analyses were conducted. These included subgroup analyses of patients with a platelet count of between $\geq$50 x 10^9/L and < 100 x 10^9/L or ≥ 100 x 10^9/L at baseline, and patients resistant and intolerant to ruxolitinib.[43, 44]

A summary of the statistical analyses in JAKARTA-2 is provided in Table 8 below.

Table 8: Summary of statistical analyses

Trial number (acronym)	NCT01523171 (JAKARTA-2)
Hypothesis objective	Fedratinib will improve spleen volume reduction in patients with myelofibrosis that have been previously treated with ruxolitinib
Statistical analysis	Spleen responses were measured using MRI/CT and continuous variables were summarised using descriptive statistics (i.e. n, mean, median, SD, min, max) A one-sided significance level of α = 0.25 was used for hypothesis testing and CIs were calculated using the two-sided 95% CI unless otherwise specified Chi-squared testing was not performed due to the early termination of the study

Sample size, power calculation	Assuming 25% of patients achieved the primary endpoint of a ≥ 35% reduction in spleen volume from baseline, 70 evaluable patients were required to provide at least 90% power to test the null hypothesis of ≤ 10% of patients achieving the primary endpoint Based on the COMFORT-I study results, ~ 60% of patients receiving ruxolitinib were non-responders. Therefore, 60% of 70 evaluable patients (i.e. 42) were required to provide 80% power to test a spleen response rate ≤ 10% for the subgroup of patients who did not reach the primary endpoint of spleen response during the ruxolitinib studies
Data management, patient withdrawals	In the original analysis, the LOCF method was used to account for patients that did not meet EOC6 due to the clinical hold In the updated analyses presented in this submission (full ITT population and reanalysis populations), LOCF was not applied. A patient without a Cycle 6 assessment was considered a non-responder The CSR provides efficacy results in ITT and PP populations with and without LOCF. Results from the ITT population without LOCF are presented in Section B.2.6 as this is considered the most robust and replicable of the datasets. Results from the PP population with LOCF are presented in the appendices
Key: CI, confidence interval; CSR, clinical study report; CT, computed tomography; EOC6, end of Cycle 6; ITT, intent-to-treat; LOCF, last observation carried forward; max, maximum; min, minimum; MRI, magnetic resonance imaging; n, number of observations; PP, per protocol; SD, standard deviation; SVR, spleen volume reduction. **Source:** Harrison et al. 2017[37] and JAKARTA-2 CSR.[43]	

All patients in JAKARTA-2 discontinued study treatment: 63 (65%) due to the early termination of the study, 18 (19%) due to AEs, six (6%) due to disease progression, three (3%) because of patient decision and seven (7%) for other reasons.[37]

Further information regarding the participant flow in JAKARTA-2 is presented in Appendix D.

B.2.5. Quality assessment of the relevant clinical effectiveness evidence

JAKARTA-2 is generally considered a high-quality study, being conducted in accordance with the ethical principles of Good Clinical Practice according to the International Council for Harmonisation guidelines.[43]

A panel of independent central readers evaluated the MRI/CT imaging scans and were blinded to reduce the potential bias in the evaluation process.[43] As this was a single-arm study, there was no risk of bias with regards to comparative evaluation. However, the single-arm design of JAKARTA-2 has the limitation of being unable to provide direct comparative evidence. Instead, evidence for the efficacy of fedratinib versus BAT, in patients with myelofibrosis who have been treated with ruxolitinib, is demonstrated by ITC (see Section B.2.9).

Potential bias may have resulted from the early termination of the fedratinib programme.[43] In particular, 65% of the patients enrolled in JAKARTA-2 were mandated to discontinue treatment due to the early termination of the study. This meant that many patients had missing data at EOC6 and additional populations and analyses were undertaken to address this limitation. This included the LOCF method in the PP population conducted in the original analyses, which provided the most optimistic results for the efficacy of fedratinib in JAKARTA-2 (see Appendix D; note these results were not reproducible by Celgene). Celgene conducted analyses in a Sensitivity Cohort to address missing data at EOC6 and demonstrated that these results were consistent with the ITT population (see Section B.2.6).

Feedback received from clinicians at an advisory board indicated that removal of intermediate-1 patients from the ITT population of JAKARTA-2 (subsequently referred to as the Int-2/high-risk population n = 81) provided evidence that is representative of patients anticipated to receive fedratinib in the UK.[18] As such, this population is considered a reliable indication of the clinical and economic benefit offered by fedratinib and is presented throughout this submission.

A complete quality assessment for the JAKARTA-2 trial, based on the NICE-recommended checklist for bias, is provided in Appendix D.

B.2.6. Clinical effectiveness results of the relevant trials

B.2.6.1 Overview

The efficacy of fedratinib in patients who have been treated with ruxolitinib has been demonstrated in JAKARTA-2 and is supported by similar efficacy in the JAK inhibitor-naïve patient population from JAKARTA.[37, 48] These results indicate that

fedratinib is efficacious in both populations, despite patients who have been treated with ruxolitinib being associated with more advanced myelofibrosis, a higher disease burden, and poorer outcomes.

The efficacy of fedratinib has been confirmed with a consistent benefit in the populations measured, including:

- The ITT population, which preserves sample size and provides the most conservative estimate
- The Int-2/high-risk population, which reflects the patients that would receive fedratinib in UK clinical practice
- The reanalysis: stringent criteria cohort, that applies more stringent criteria of relapse/refractory than the ITT population
- Reanalysis: sensitivity cohort, adjusting for the impact of the clinical hold

In the ITT population of JAKARTA-2, 31% of patients achieved the primary outcome of spleen response rate defined as ≥35% SVR at EOC6.[46] Similarly, 27% of patients achieved the key secondary outcome of symptom response rate defined as ≥50% reduction in TSS at EOC6. An overview of the results for major endpoints in the key populations from JAKARTA-2 and JAKARTA is provided in Table 9.

Full trial results for the ITT JAKARTA-2 population are presented in subsequent sections. Additionally, results for SVR, TSS and SVR or TSS endpoints in the other key populations are presented to demonstrate consistent benefit.

Supporting results from JAKARTA and the PP population with LOCF from JAKARTA-2 are provided in Appendix D.

Table 9: Overview of fedratinib efficacy, JAKARTA-2 and JAKARTA

Endpoint	Measure	JAKARTA-2 (Phase II, previously treated with ruxolitinib)				JAKARTA (Phase III, ruxolitinib naïve)		
		ITT population	Int-2 and high-risk patients[a]	Reanalysis: Stringent Criteria Cohort[b]	Reanalysis: Sensitivity Cohort[c]	ITT population		
		Fedratinib 400 mg (n=97)	Fedratinib 400 mg (n=81)	Fedratinib 400 mg (n=79)	Fedratinib 400 mg (n=66)	Placebo (n=96)	Fedratinib 400 mg (n=96)	Fedratinib 500 mg (n=97)
Spleen response rate	≥35% SVR at EOC6, % (95% CI)	31% (22, 41)	[redacted]	30% (21, 42)	36% (25, 49)	1% (0, 3)	47% (37, 57)	50% (40, 59)
Symptom response rate[d]	≥50% reduction in TSS at EOC6, % (95% CI)	25% (17, 35)	[redacted]	27% (17, 39)	32% (21, 45)	7% (2, 13)	36% (26, 46)	34% (24, 44)
Overall survival	12-month OS rate, % (95% CI)	[redacted]	[redacted]	NA	[redacted]	[redacted]	[redacted]	[redacted]
	HR versus placebo (95% CI)	NE	NE	NE	NE	NA	[redacted]	[redacted]

Key: CI, confidence interval; CSR, clinical study report; EOC6, end of Cycle 6; HR, hazard ratio; int, intermediate; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; NA, not assessed; NE, not estimable; OS, overall survival; TSS, total symptom score.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold; [d], this outcome was assessed in the MF-SAF population which was defined as patients with evaluable baseline and ≥1 post-baseline MF-SAF assessment. For JAKARTA-2, this includes 90 of the ITT patients, 74 of the Stringent Criteria Cohort patients, and 62 of the Sensitivity Cohort patients. For JAKARTA, this includes 91 patients in the fedratinib groups and 85 patients in the placebo group; [e], symptom response rate in the Int-2/high-risk subgroup did not apply evaluable baseline and ≥1 post-baseline MF-SAF assessment criteria.
Source: Harrison et al. 2020,[46] Pardanani et al. 2015,[48] JAKARTA-2 CSR[43] and data on file.[49, 50]

B.2.6.2 Primary outcome: spleen response rate (≥ 35% SVR) at EOC6

Treatment with fedratinib is associated with a significant spleen response rate, with 31% of patients achieving ≥ 35% SVR at EOC6, which the International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consider an appropriate threshold for response in patients with myelofibrosis.[46, 51] These results were consistent in the Int-2/high-risk subpopulation, for which [redacted]% (95% CI: [redacted]%, [redacted]%) of patients achieved ≥ 35% SVR at EOC6.[50]

Results from the reanalysis, applying more stringent criteria of ruxolitinib relapse and intolerance to the ITT population, found results were concordant with the ITT population; with 30% of Stringent Criteria Cohort patients demonstrating ≥ 35% SVR at EOC6 (95% CI: 21%, 42%).[46] When removing patients who were directly impacted by the clinical hold (i.e. the Sensitivity Cohort), 36% of patients demonstrated SVR at EOC6 (95% CI: 25%, 49%).

Table 10: Spleen response rates at EOC6 (≥ 35% SVR; JAKARTA-2)

≥ 35% SVR at EOC6, n, % (95% CI)	Fedratinib 400 mg			
	ITT population (n = 97)	Int-2/high-risk patients[a] (n = 81)	Reanalysis: Stringent Criteria Cohort[b] (n = 79)	Reanalysis: Sensitivity Cohort[c] (n = 66)
	30, 31% (22, 41)	[redacted]	24, 30% (21, 42)	24, 36% (25, 49)

Key: CI, confidence interval; EOC6, end of Cycle 6; int, intermediate; ITT, intent-to-treat; SVR, spleen volume reduction.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold.
Source: Harrison et al. 2020[46], and data on file.[50]

B.2.6.3 Secondary outcome measures

B.2.6.3.1 Spleen response rate (≥ 35% SVR) at the EOC3

Treatment with fedratinib is associated with almost half of patients achieving ≥ 35% SVR at EOC3, which the IWG-MRT and ELN regard as a lasting benefit qualifying a response.[43, 51]

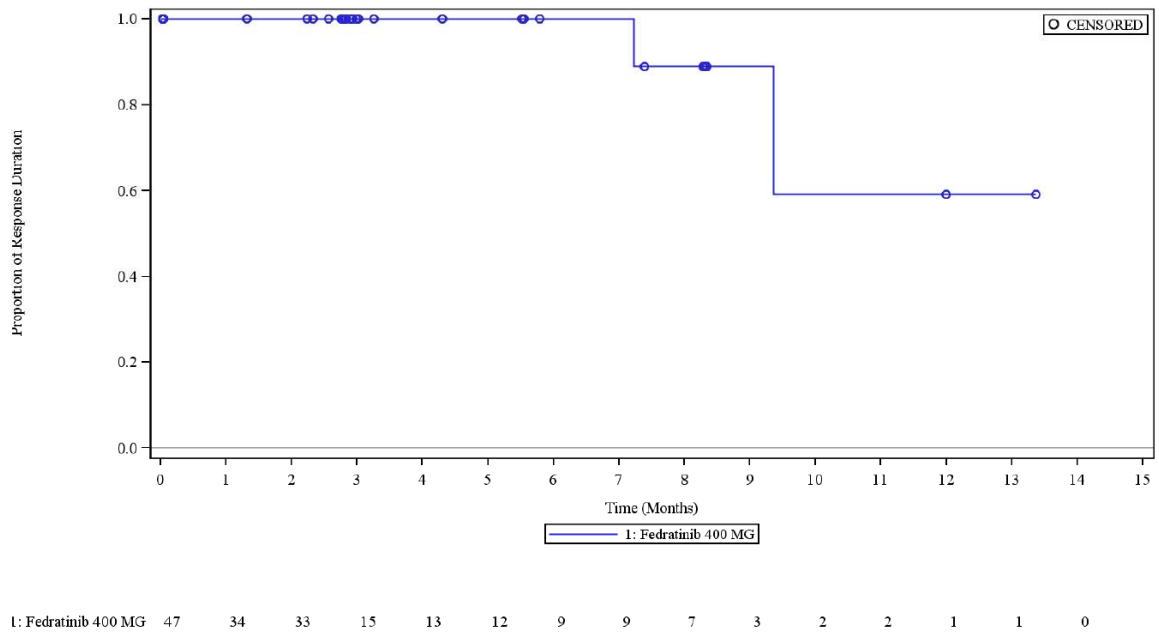
The proportion of patients with ≥ 35% SVR at EOC3 were 40% (95% CI: 30%, 51%) in the ITT population, 43% (95% CI: 32%, 55%) in the Stringent Criteria Cohort, and 41% (95% CI: 29%, 54%) in the Sensitivity Cohort.[46]

B.2.6.3.2 Duration of spleen response

Treatment with fedratinib is associated with the majority of patients achieving a duration of response longer than 9 months, although this outcome measure required extensive censoring due to early termination.[46]

For the duration of response analysis, responders were all patients who at any time achieved ≥ 35% SVR from baseline: this included 47 patients in JAKARTA-2 (Figure 7).[46] Based on Kaplan–Meier (KM) estimates, only 25% of patients had a duration of response of less than 9.4 months and the median duration was not reached (NR). Median spleen volume response duration was also NR (95% CI: 7.2 months, NR) in both the Stringent Criteria Cohort (n = 41 responders) and the Sensitivity Cohort (n = 34 responders).

Figure 7: Kaplan–Meier plot of duration of spleen response, ≥ 35 SVR at any time on study treatment (JAKARTA-2, ITT population)



Key: ITT, intent-to-treat; SVR, spleen volume reduction.
Notes: patients at risk are shown along the horizontal axis. The duration of spleen response was calculated from the first date of spleen response (i.e. ≥ 35% SVR from baseline) to the first date of

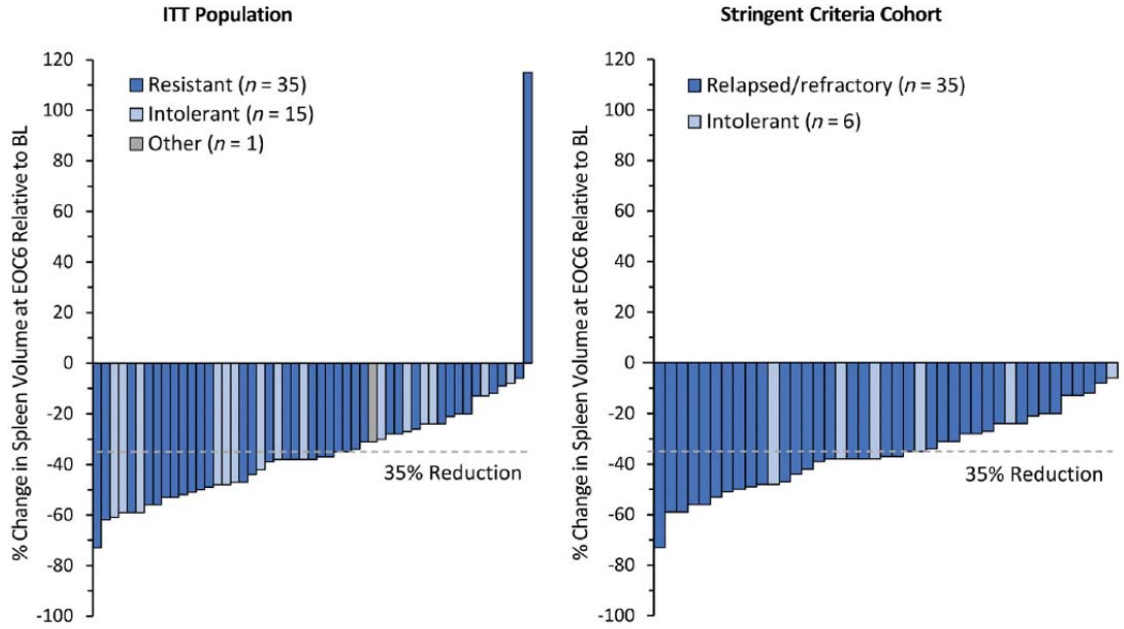
disease progression (i.e. ≥ 25% spleen volume increase from baseline) or death, whichever was earlier.
Source: Harrison et al. 2020.[46]

B.2.6.3.3 Percent change of spleen volume at EOC3 and EOC6

Treatment with fedratinib is associated with the majority of patients achieving a reduction in spleen volume, with an average reduction of one-third.[46] In the ITT population, the median percentage changes in spleen volume were [redacted]% at EOC3 (range: [redacted]) and -38.0% at EOC6 (range: -73, -115).[43, 46]

When considering individual changes in spleen volume for patients with measurements at baseline and EOC6, all patients except one in the ITT population showed a reduction in volume.[46] In the Stringent Criteria Cohort all patients showed a SVR (Figure 8).

Figure 8: Individual changes in spleen volume from baseline to EOC6 (JAKARTA-2, ITT and Stringent Criteria Cohort)



Key: BL, baseline; EOC6, end of Cycle 6; ITT, intent-to-treat.
Source: Harrison et al. 2020.[46]

B.2.6.3.4 Spleen response rate by palpation at EOC3 and EOC6

Spleen response rate by palpation was defined as the proportion of patients with ≥ 50% reduction in spleen size.[43] In the ITT population of JAKARTA-2, treatment with fedratinib was associated with considerable reductions in spleen size, with almost one third of patients treated achieving at ≥ 50% reduction in size, which the IWG-MRT and ELN consider a clinically meaningful response in patients with myelofibrosis.[43, 51] In the ITT population, the proportion of patients with ≥ 50% reduction in spleen size were [redacted]% at EOC3 and 31% at EOC6 (Table 11).[43, 46]

Of note, the patients that demonstrated ≥ 35% SVR at EOC6 were the same patients who demonstrated ≥ 50% reduction in spleen size at EOC6. This supports previous literature that suggests these outcomes are highly consistent or equivalent.[52-54]

Table 11: Spleen response rate by palpation (≥ 50% reduction in spleen size) at EOC3 and EOC6 (JAKARTA-2, ITT population)

	Fedratinib 400 mg (N=97)
EOC3	
n (%)	[redacted]
95% CI	[redacted]
EOC6	
n (%)	30 (31%)
95% CI	[redacted]

Key: CI, confidence interval; EOC3, end of Cycle 3; EOC6, end of Cycle 6; ITT, intent-to-treat.
Notes: Spleen size was measured by palpitation (i.e. length in cm)
Source: JAKARTA-2 CSR[43], and Harrison 2020.[46]

Results in the Stringent Criteria Cohort and Sensitivity Cohort were consistent with the ITT population, with reduction in spleen size of ≥ 50% at EOC6 observed in 30 (31%) and 24 (36%) patients, respectively.[46]

B.2.6.3.5 Symptom response rate (≥ 50% reduction in TSS) at EOC6

The analyses of symptom response rate were performed using the MF-SAF Analysis Population, defined as patients with an evaluable baseline assessment of modified MF-SAF TSS, and at least one post-baseline evaluable assessment.[43] Symptom

response rates were defined as the proportion of patients with ≥50% reduction in TSS from baseline to EOC6.

Treatment with fedratinib was associated with considerable symptom relief, with most evaluable patients having demonstrated an improvement in TSS and more than a quarter achieving the clinically meaningful threshold for response of ≥ 50% reduction.[46, 51] The proportion of patients in the MF-SAF Analysis Population with a ≥ 50% reduction in TSS at EOC6 was 27% (95% CI: 18%, 37%). Among patients with evaluable TSS data at baseline and EOC6, 82% reported some decrease in symptom severity with fedratinib.

Symptom response rates in the Stringent Criteria and Sensitivity Cohorts supported results for the ITT Population.[46] At EOC6, symptom response rates were 27% (95% CI: 17, 39) and 32% (95% CI: 21, 45), respectively (Table 12).

In order to derive the most conservative plausible estimate, and to ensure comparability with reporting in other trials, the economic modelling for fedratinib calls upon results for symptom response rate in the ITT population. In this population, the proportion of patients with ≥ 50% reduction in TSS at EOC6 was [redacted]% ([redacted]/97).[50] The results for the Int-2/high-risk subgroup of patients were consistent with the ITT population with [redacted]% (95% CI: [redacted]) achieving ≥ 50% reduction in TSS at EOC6 ([redacted]/81).

Table 12: Symptom response rates at EOC6 (≥ 50% TSS; JAKARTA-2)

≥ 50% reduction in TSS at EOC6	**Fedratinib 400 mg**			
	All enrolled	**Int-2/high-risk patients[a]**	**Reanalysis: Stringent Criteria Cohort[b]**	**Reanalysis: Sensitivity Cohort[c]**
MF-SAF, N[d]	90	NA	74	62
% (95% CI)	27% (18, 37)	NA	27% (17, 39)	32% (21, 45)

≥ 50% reduction in TSS at EOC6	Fedratinib 400 mg			
	All enrolled	**Int-2/high-risk patients[a]**	**Reanalysis: Stringent Criteria Cohort[b]**	**Reanalysis: Sensitivity Cohort[c]**
ITT, N	97	81	NA	NA
n, % (95% CI)	[redacted]	[redacted]	NA	NA

Key: CI, confidence interval; CSR, clinical study report; EOC6, end of Cycle 6; int, intermediate; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; NA, not assessed.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold; [d], includes patients with evaluable baseline and ≥ 1 post-baseline MF-SAF assessment.
Source: Harrison et al. 2020[46], and data on file.[50]

Total symptom score by key symptoms

All key symptoms assessed in the MF-SAF Analysis Population in JAKARTA-2 showed an improvement at EOC6 in half of the evaluable patients, with median percent changes of:[46]

- -83% in pain under ribs on left side
- -76% in night sweats
- -51% in early satiety
- -46% in abdominal discomfort
- -44% in pruritus
- -222% in bone or muscle pain

These results indicate that treatment with fedratinib is associated with relief of many of the constitutional symptoms of myelofibrosis.

B.2.6.4 Key exploratory outcome measures

B.2.6.4.1 Spleen or symptom response rate at EOC6

A combined endpoint of spleen or symptom response was strongly recommended as a modelling input by experts at an advisory board, with the rationale that this outcome would be reflective of UK clinical practice given that the two track together.[18] Spleen or symptom response rate is defined as the number of patients achieving either ≥ 35% SVR or ≥ 50% reduction in TSS.

At EOC6 in JAKARTA-2, treatment with fedratinib was associated with almost half of patients achieving a spleen or symptom response rate, with generally consistent results in ITT, Int-2/high-risk, and reanalysis cohorts (Table 13).[50]

Table 13: Spleen or symptom response rates at EOC6 (≥ 35% SVR or ≥ 50% reduction in TSS; JAKARTA-2)

≥ 35% SVR or ≥ 50% reduction in TSS at EOC6, n, % (95% CI)	Fedratinib 400 mg			
	ITT population (n=97)	Int-2/high-risk patients[a] (n=81)	Reanalysis: Stringent Criteria Cohort[b] (n=79)	Reanalysis: Sensitivity Cohort[c] (n=66)
	[redacted]	[redacted]	[redacted]	[redacted]

Key: CI, confidence interval; CSR, clinical study report; EOC6, end of Cycle 6; int, intermediate; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; NA, not assessed.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold; [d], includes patients with evaluable baseline and ≥1 post-baseline MF-SAF assessment.
Source: Data on file.[50]

B.2.6.4.2 Overall survival

Although promising, OS data for JAKARTA-2 are immature and heavily censored owing to early study termination (see Section B.2.4). At the time of the final analysis there were a total of [redacted] deaths; [redacted] deaths occurred whilst on-treatment and [redacted] deaths occurred more than 30-days post treatment.[43] The proportion of patients alive at 12 months was [redacted]% (Figure 9).[49]

Figure 9: Kaplan–Meier estimates of OS (JAKARTA-2, ITT population)

Key: ITT, intent-to-treat; NC, not calculable; OS, overall survival.
Notes: OS is defined as the time interval from the date of first dose to the date of death due to any cause. In the absence of the confirmation of death, OS is censored at the last date patient was known to be alive.
Source: Data on file.[49]

OS estimates for the Int-2/high-risk disease population were consistent with the ITT population, with [redacted]% patients alive at 12 months.[50]

Figure 10: Kaplan–Meier estimates of OS (JAKARTA-2, Int-2/high-risk population)

Key: Int, intermediate; OS, overall survival.
Notes: OS is defined as the time interval from the date of first dose to the date of death due to any cause. In the absence of the confirmation of death, OS is censored at the last date patient was known to be alive.
Source: Data on file.[50]

JAKARTA demonstrated an OS benefit for fedratinib 400 mg versus placebo, with [redacted]% of patients alive at 12 months (OS hazard ratio [HR] [redacted]; 95% CI: [redacted], [redacted]; p=[redacted]).[49] This implies that there may also be an OS benefit for fedratinib in patients treated with ruxolitinib; although, this should be interpreted with caution given the differences in patient populations.

Feedback received from clinicians indicates there is clinical plausibility for the use of SVR as a surrogate marker for survival.[18] The SVR outcomes observed in JAKARTA and JAKARTA-2, taken together with the OS benefit versus placebo observed in JAKARTA, further support the idea that fedratinib offers a survival benefit to patients treated with ruxolitinib.

B.2.6.4.3 EORTC QLQ-C30

EORTC QLQ-C30 analyses were undertaken in the EORTC QLQ-C30 analysis population (n = 90), defined as all treated patients who had a baseline and ≥ 1 post-

baseline assessment of the QLQ-C30 questionnaire.[43] Completion rates for patients in the ITT population for each cycle were high, ranging from [redacted]% to [redacted]% for all cycles.

Treatment with fedratinib was associated with improvements in HRQoL, with [redacted] of evaluable patients having demonstrated post-baseline improvements in global quality of life (QoL), physical functioning, fatigue, pain, insomnia and appetite loss.[43] For all other functional and symptom domains, HRQoL was maintained over the six-cycle treatment.

The QLQ-C30 is a widely used cancer-specific instrument made up of functional domains (for which a higher score indicates better HRQoL) and symptom domains (for which a lower score indicates a better HRQoL).[55] At EOC6, mean changes from baseline in QLQ-C30 functional domain scores were:[43]

- GHS QoL – [redacted]
- Physical functioning domain – [redacted]
- Role functioning domain – [redacted]
- Social functioning domain – [redacted]

For symptom domain scores, considerable improvements in mean change in QLQ-C30 score from baseline to EOC6 were observed for appetite loss ([redacted]), insomnia ([redacted]), dyspnoea ([redacted]), financial difficulties ([redacted]), fatigue ([redacted]) and pain ([redacted]).[43]

The mean changes from baseline in EORTC QLQ-C30 functional and symptom scores are presented in Figure 11 and Figure 12, respectively.

Figure 11: Mean change from baseline in QLQ-C30 functional scores (JAKARTA-2, EORTC QLQ-C30 Analysis Population)

Key: EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Core Quality of Life.
Source: JAKARTA-2 CSR.[43]

Figure 12: Mean change from baseline in QLQ-C30 symptom scores (JAKARTA-2, EORTC QLQ-C30 Analysis Population)

Key: EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Core Quality of Life.
Source: JAKARTA-2 CSR.[43]

B.2.7. Subgroup analysis

Subgroup analyses were carried out to determine the treatment effect of fedratinib on clinically important subpopulations. These analyses included spleen response rate (≥ 35% SVR) and symptom response rate (≥ 50% reduction in TSS) by baseline demographic and disease characteristics, as well as in subgroups of patients with a platelet count of between ≥ 50 x 10^9/L and < 100 x 10^9/L or ≥ 100 x 10^9/L at baseline and patients resistant or intolerant to ruxolitinib.[43]

Overall, results of the subgroup analyses of spleen response rate and symptom response rate were consistent across baseline demographic and disease characteristics subgroups, supporting the robustness of the results of the primary analysis (see Appendix E).[43]

Irrespective of the baseline platelet count at baseline, fedratinib showed clinical benefit in terms of spleen response rate and symptom response rate (Table 14).[43, 46]

Table 14: Efficacy of fedratinib 400 mg by platelet count at baseline (JAKARTA-2)

	Platelet count at baseline	
	≥ 50 x 10^9/L to < 100 x 10^9/L	**≥ 100 x 10^9/L**
≥ 35% SVR at EOC6[a]		
ITT population, n/N	[redacted]/33	[redacted]/64
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC3[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC6[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]

Key: CI, confidence interval; CT, computed tomography; EOC3, end of Cycle 3; EOC6, end of Cycle 6; MF-SAF, myelofibrosis symptom assessment form; MRI, magnetic resonance imaging; SVR, spleen volume reduction; TSS, total symptom response.
Notes: [a], spleen volume was measured by MRI/CT scan and reviewed in a blinded fashion by a central imaging laboratory; [b], CI estimated using Clopper–Pearson Exact method; [c], TSS was defined as the sum of the daily average score of the six-item measures in a week: night sweats, pruritus, abdominal discomfort, early satiety, pain under ribs on left side and bone or muscle pain. For this analysis, patients with a baseline TSS equal to 0 are excluded (due to no place for symptom reduction). Patients with a missing TSS at the EOC6 were considered as non-responders.
Source: Harrison 2020[46], and JAKARTA-2 CSR.[43]

Similarly, fedratinib showed clinical benefit in terms of spleen response rate and symptom response rate, irrespective of ruxolitinib status (resistant versus intolerant) at baseline (Table 15).[43, 46]

Table 15: Efficacy of fedratinib 400 mg in patients resistant or intolerant to ruxolitinib at baseline (JAKARTA-2, ITT population)

	Resistant	Intolerant
≥ 35% SVR at EOC6[a]		
ITT population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC3[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]

	Resistant	Intolerant
≥ 50% reduction in TSS at EOC6[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]

Key: CI, confidence interval; EOC3, end of Cycle 3; EOC6, end of Cycle 6; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; SVR, spleen volume reduction; TSS, total symptom score.
Notes: Investigators' assessments of patients resistant or intolerant to ruxolitinib. One patient was neither resistant nor intolerant per investigator's assessment and was categorised under 'other: lack of efficacy'. [a], Spleen volume was measured by MRI/CT scan and reviewed in a blinded fashion by a central imaging laboratory; [b], CI estimated using Clopper–Pearson Exact method; [c], TSS was defined as the sum of the daily average score of the six-item measures in a week: night sweats, pruritus, abdominal discomfort, early satiety, pain under ribs on left side and bone or muscle pain. For this analysis, patients with a baseline TSS equal to 0 are excluded (due to no place for symptom reduction). Patients with a missing TSS at the EOC6 were considered as non-responders.
Source: Harrison 2020[46], and JAKARTA-2 CSR.[43]

B.2.8. Meta-analysis

As a single study (JAKARTA-2) provides data for fedratinib in patients treated with ruxolitinib, meta-analysis of intervention studies is not required.

An ITC has been conducted to demonstrate the comparative efficacy and safety of fedratinib versus BAT, and is described in detail in Section B.2.9.

B.2.9. Indirect treatment comparison

B.2.9.1 Background

The comparative efficacy and safety of fedratinib versus BAT cannot be directly inferred from the JAKARTA-2 trial as it is a single-arm study; therefore, comparative evidence needs to be calculated using an ITC.

B.2.9.2 Methods

An SLR was conducted to identify evidence of relevance to the efficacy and safety of treatments for myelofibrosis.[56] The SLR identified three studies that investigated either fedratinib or BAT in a patient population that had received prior JAK-inhibitor treatment; JAKARTA-2, PERSIST-2 and SIMPLIFY-2. These trials were included as they investigated SVR and/or TSS reduction and could therefore be compared with

evidence for fedratinib from JAKARTA-2. See Appendix D for full details of the methods and results of the SLR.

The ITT populations from each of these trials represents the most appropriate populations for comparative purposes. Of note, the inclusion of intermediate-1 patients from JAKARTA-2 provides the most conservative estimate of efficacy and therefore is not thought to bias in favour of fedratinib.

Given JAKARTA-2 is a single-arm trial, the two methods that were explored to perform an unanchored indirect comparison of fedratinib with BAT were matching-adjusted indirect comparison (MAIC) and simulated treatment comparison (STC) methods. Both methods were discussed in the NICE Decision Support Unit and Technical Support Document 18. The Technical Support Document states that there is little in the literature to suggest one methodology is superior to the other.[57] An unanchored MAIC or STC assumes that all effect modifiers and prognostic factors are accounted for. This assumption is largely considered impossible to meet.

As no baseline characteristics specific to the JAK-inhibitor exposed population were available for the PERSIST-2 study, only a naïve comparison between fedratinib in JAKARTA-2 and BAT in PERSIST-2 was feasible.[56]

Baseline characteristics were available for SIMPLIFY-2. As such, MAIC and STC analyses, controlling for baseline characteristics identified as being both prognostic and imbalanced between data sources, were explored for comparisons between fedratinib and BAT. Given JAKARTA-2 is a single-arm study it was not possible to use the patient-level data to identify treatment effect modifiers and no information on treatment effect modifiers for this population was found in the literature. See Appendix D for full details of baseline characteristics, methods and statistical analyses used in the ITC.

A summary of the trials used to carry out the indirect treatment comparisons is provided in Table 16.

Table 16: Summary of the trials used in the indirect treatment comparison

	JAKARTA-2	PERSIST-2	SIMPLIFY-2 [a]
Phase	II	III	III
Design	Single-arm	RCT	RCT
Method of blinding	Open-label	Open-label	Open-label
Intervention (N)	Fedratinib 400 mg, once daily (starting dose) (97 [ITT])	Pacritinib 400 mg, once daily (75 [ITT]) and pacritinib 200 mg, twice daily (74 [ITT])	Momelotinib 200 mg once daily (104 [ITT])
Comparator	NA	BAT (72 [ITT efficacy population]): • Ruxolitinib (45%) • Watch and wait (19%) • Hydroxycarbamide (hydroxyurea)(19 %) • Prednisone (13%) • Danazol (5%) • Thalidomide (3%) • Decitabine (2%) • Interferon-alpha (2%)	BAT (52 [ITT]): • Ruxolitinib (89%) • Hydroxycarbamide (hydroxyurea) (23%) • Corticosteroids (12%)
Location	Multicentre	Multicentre	Multicentre
Method of randomisation	NA	1:1:1 ratio stratified by geographic region, risk category and rebound platelet count	2:1 stratified by transfusion dependence and by baseline TSS
Crossover	NA	After Week 24 or progression of splenomegaly before Week 24	After completion of the randomized phase (24 weeks), all subjects were eligible to receive momelotinib in an extended treatment phase
Key inclusion/exclusion criteria			
Prior JAK inhibitor treatment	Prior ruxolitinib (≥14 days of exposure or <14 days if patients discontinued ruxolitinib due to intolerability or allergy)	Prior treatment with one or two other JAK-inhibitors was allowed, patients could be JAK-inhibitor naïve:	Currently or previously treated with ruxolitinib (≥28 days) and either: • RBC transfusion needed while on ruxolitinib

	JAKARTA-2	PERSIST-2	SIMPLIFY-2 [a]
		• 33 patients had prior ruxolitinib (45.8%) • 39 patients were ruxolitinib naïve (54.2%)	• Dose adjustment of ruxolitinib to <20 mg twice daily and Grade 3 thrombocytopenia/ anaemia/hematom a
Platelet count	≥50 x 10^9/L	≤100 x 10^9/L	There were no inclusion/exclusion criteria for platelet count at baseline
Diagnosis	PMF, PPV-MF, PET-MF	PMF, PPV-MF, PET-MF	PMF, PPV-MF, PET-MF
DIPSS[b]	• Intermediate-1 with symptoms • Intermediate-2 • High-risk	• Intermediate-1 • Intermediate-2 • High-risk	• Intermediate-1 with symptomatic splenomegaly/ hepatomegaly • Intermediate-2 • High-risk

Key: BAT, best available therapy; DIPSS, Dynamic International Prognostic Scoring System; ITC, indirect treatment comparison; ITT, intent-to-treat; L, litre; N, number of subjects; NA, not applicable; PET-MF, post-essential thrombocytopenia myelofibrosis; PMF, primary myelofibrosis; PPV-MF, post-polycythaemia vera myelofibrosis; RBC, red blood cell; RCT, randomised controlled trial; TSS, total symptom score.
Notes: [a], only the most frequent treatments received were reported; the percentages in this table do not sum to 100% as patients could have received more than one therapy; [b], DIPSS score calculation: 1 point for each of the following criteria: age >65 years, white cell count ≥25 x 10^9/L, haemoglobin < 10 g/dL, peripheral blood blasts ≥1%, constitutional symptoms (weight loss and/or unexplained fever or excessive sweats).
Source: ITC report.[56]

One of the key differences in study inclusion/exclusion criteria was platelet count at baseline; with JAKARTA-2 only including patients with a platelet count of ≥50 x 10^9/L, PERSIST-2 including patients with ≤100 x 10^9/L and SIMPLIFY-2 not applying a limit (Table 16).[56] To account for this, the naïve comparison of PERSIST-2 and JAKARTA-2 was conducted on the subgroup of patients in JAKARTA-2 with a platelet count <100 x 10^9/L (see Section B.2.9.5.2).

B.2.9.3 Results

All three studies collect the following two efficacy outcomes of interest to the ITC:

- The proportion of patients achieving ≥35% SVR at 24 weeks from baseline
- The proportion of patients achieving ≥50% TSS reduction at 24 weeks from baseline

PERSIST-2 informed a naïve ITC comparing fedratinib to BAT in patients with a platelet count <100 x 10^9/L.[56] In this comparison, treatment with fedratinib was associated with a greater proportion of patients achieving ≥35% SVR ([redacted]% greater; 95% CI: [redacted]) and a greater proportion of patients achieving ≥50% reduction in TSS ([redacted]% greater; 95% CI: [redacted]) (Table 17).

Table 17: Summary of comparisons to the PERSIST-2 evidence

Comparison made	**Data used to make the comparison**	
	JAKARTA-2: fedratinib 400 mg (N=33)[a]	**PERSIST-2: BAT (N=33)[b]**
Proportion of ITT subjects with platelet count <100 x 10^9/L achieving ≥35% SVR from baseline to Week 24/EOC6, n (%)[c]	[redacted] ([redacted]%)	1 (3%)
Naïve ITC for ITT subjects with platelet count <100 x 10^9/L achieving ≥35% SVR from baseline to Week 24/EOC6, RD (95% CI)[c]	[redacted]% ([redacted])	
Proportion of ITT subjects with platelet count <100 x 10^9/L achieving ≥50% TSS reduction from baseline to Week 24/EOC6, n (%)	[redacted] ([redacted]%)	5 (15%)
Naïve ITC for ITT subjects with platelet count <100 x 10^9/L achieving ≥50% TSS reduction from baseline to Week 24/EOC6, RD (95% CI) [c]	[redacted]% ([redacted])	

Key: BAT, best available therapy; EOC, end of cycle; ITC, indirect treatment comparison; L, litre; N, total number of subjects; RD, risk difference; SVR, spleen volume reduction; TSS, total symptom score.
Notes: [a], denominator refers to the number of patients from JAKARTA-2 with platelet count of <100 x 10^9/L at baseline; [b], denominator refers to patients from PERSIST-2 that had previously been treated with ruxolitinib; RD calculated by subtracting the proportion of BAT responders from the proportion of fedratinib responders; [c], this row indicates absolute responses and is not an ITC.
Source: ITC report.[56]

Adjusted analyses using PERSIST-2 was not possible due to the paucity of publicly available baseline characteristics for the ruxolitinib exposed population; therefore, they were conducted using the JAKARTA-2 and SIMPLIFY-2 studies only.[56] These analyses were conducted in the ITT population for JAKARTA-2, presented below. Results were consistent in an ITC conducted using the Sensitivity Cohort from JAKARTA-2.

Identification of imbalanced prognostic factors to adjust for in the ITC was performed as follows[56]:

- The variable was identified as imbalanced across the JAKARTA-2 study and the BAT arm of the SIMPLIFY-2 study based on an external haematologist identifying the imbalance as clinically meaningful
- The variable was identified as being an important prognostic factor based on univariable and multivariable analyses performed with the JAKARTA-2 patient-level data

Variables fulfilling both criteria for SVR were ECOG PS and transfusion dependence, and variables fulfilling both criteria for TSS reduction were ECOG PS and DIPSS.[56] A full list of the variables explored is described in Appendix L.

The ITC results adjusting for prognostic variables are presented in Table 18 and Table 19 below.[56] The reweighted baseline characteristics for each of the adjusted analyses are presented in Appendix L.

Table 18: Naïve and adjusted ITC results for SVR (JAKARTA-2 and SIMPLIFY-2)

Method	**Variables included in adjustment[b]**	**JAKARTA-2 (fedratinib 400 mg; N=97)**	**SIMPLIFY-2 (BAT; N=52)**
Naïve ITC	• NA	30.9% (n=30)	5.8% (n=3)
		RD[c] (95% CI): 25.2% (14, 36.3)	
MAIC	• ECOG PS	[redacted]% (CI: [redacted])[a]	5.8% (n=3)

Method	Variables included in adjustment[b]	JAKARTA-2 (fedratinib 400 mg; N=97)	SIMPLIFY-2 (BAT; N=52)
			RD[c] (95% CI): [redacted][a]
STC	• ECOG PS	[redacted] (CI: [redacted])	5.8% (n=3)
			RD[c] (95% CI): [redacted]
MAIC	• ECOG PS • Transfusion dependence	[redacted] (CI: [redacted])[a]	5.8% (n=3)
			RD[c] (95% CI): [redacted][a]

Key: BAT, best available therapy; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ESS, effective sample size; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; n, number of responders; N, total number of patients; NA, not applicable; RD, risk difference; STC, simulated treatment comparison; SVR, spleen volume reduction.
Note: [a], bootstrap percentile CI (based on 10,000 samples); [b], ESS of JAKARTA-2 population after matching on ECOG PS was 91.7 (94.5% of original sample size) and after matching on ECOG PS and transfusion dependence was 34.4 (35.5% of original sample size); [c] RD calculated by subtracting the proportion of BAT responders from the proportion of fedratinib responders.
Source: ITC report.[56]

For the MAIC of SVR, the matching procedure led to a relatively small effective sample size (ESS) for the JAKARTA-2 population (ESS was 34.4, 35.5% of the original sample size). As a small ESS is an indication that the weights are highly variable due to a lack of population overlap, and that the estimate may be unstable, additional analyses were performed with adjustment for ECOG PS only (in this case the ESS was 91.7).

For the STC of SVR, results are presented with adjustment for ECOG PS only. The adjustment for ECOG PS and transfusion dependence resulted in a logistic regression model that had a very large standard error for the transfusion dependence coefficient (standard error was 1,722.4 for the transfusion dependence coefficient compared with a standard error of 0.63 for the ECOG PS coefficient). The high standard error was likely due to transfusion dependence being a perfect predictor of the outcome and, therefore, the model struggled to converge. This is a problem that is referred to as complete separation.[58, 59]

When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had a 25.2% (95% CI: 14, 36.3) greater proportion of patients with ≥35% SVR compared with BAT.[56] After adjustment for baseline ECOG PS the difference in the proportion of patients with ≥35% SVR compared with BAT increased slightly; fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥35% SVR compared with BAT. After adjustment for baseline ECOG PS and transfusion dependence, fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥35% SVR compared with BAT. However, the results with adjustment for ECOG PS and transfusion dependence should be interpreted with caution given the relatively small effective sample size.

Table 19: Naïve and adjusted ITC results for TSS reduction (JAKARTA-2 and SIMPLIFY-2)

Method	**Variables included in adjustment[b]**	**JAKARTA-2 (400 mg fedratinib; N=97)**	**SIMPLIFY-2 (BAT; N=51)**
Naïve ITC	• NA	[redacted]% (n=[redacted])	5.9% (n=3)
		RD[C] (95% CI): [redacted]% ([redacted])	
MAIC	• ECOG PS • DIPSS	[redacted]% ([redacted])[a]	5.9% (n=3)
		RD (95% CI): [redacted]% ([redacted])	
STC	• ECOG PS • DIPSS	[redacted]% ([redacted])	5.9% (n=3)
		RD (95% CI): [redacted]% ([redacted])	

Key: BAT, best available therapy; CI, confidence interval, DIPSS, Dynamic International Prognostic Scoring System; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ESS, effective sample size; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; n, number of responders; N, total number of subjects; NA, not applicable; RD, risk difference; STC, simulated treatment comparison; TSS, total symptom score.
Note: [a], bootstrap percentile CI (based on 10,000 samples); [b], ESS of JAKARTA-2 population after matching on ECOG PS and DIPSS was 81.6 (84.2% of original sample size); [c], RD calculated by subtracting the proportion of BAT responders from the proportion of fedratinib responders.
Source: ITC report.[56]

For both the naïve analyses and the MAIC, fedratinib 400 mg consistently led to a greater proportion of patients achieving ≥50% reduction in TSS compared with BAT.[56] When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had an [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥50% TSS reduction compared with BAT. The MAIC, which adjusted for ECOG PS and DIPSS, showed that fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥50% TSS reduction compared with BAT. Similarly, a [redacted]% (95% CI: [redacted]) difference was observed using the STC methodology.

The feasibility of conducting an OS MAIC for these trials was assessed. The minimum criteria for investigating an MAIC of OS with BAT versus fedratinib are:

- Reports KM data for OS in the appropriate population
- Reports the baseline characteristics of the population observed

B.2.9.3.1 Comparative safety

With the exception of AEs leading to treatment discontinuation, the overall summary of safety in the fedratinib arm is acceptable in light of the overall summary of safety in the BAT arms of PERSIST-2 and SIMPLIFY-2 (Table 20).[56]

Table 20: Summary of treatment emergent AEs reported for JAKARTA-2 and BAT arms of PERSIST-2 and SIMPLIFY-2

	JAKARTA-2: fedratinib 400 mg (N=97)	**PERSIST-2: BAT (N=98 [Safety population])**	**SIMPLIFY-2: BAT (N=52)**
n (%) of patients with at least one AE	97 (100%)	87 (89%)	46 (89%)
n (%) of patients with at least one Grade 3 or 4 AE	61 (62.9%)	48 (49%)	NR
n (%) of patients with at least one SAE	33 (34.0%)	30 (31%)	12 (23%)
n (%) of patients who discontinued treatment due to AEs	19 (19.6%)	4 (4%)	1 (2%)
n (%) of patients with AEs leading to death	7 (7.2%)	9 (9%)[a]	4 (8%)

	JAKARTA-2: fedratinib 400 mg (N=97)	PERSIST-2: BAT (N=98 [Safety population])	SIMPLIFY-2: BAT (N=52)
n (%) of patients with dose interruption for at least 7 consecutive days	25 (25.8%)	10 (10%)[b]	NR
n (%) of patients with dose reduction	38 (39.2%)	7 (7%)	NR

Key: AE, adverse event; BAT, best available therapy; N, number of patients; NR, not reported; SAE, serious adverse event.
Note: [a], percent is given for N=100; [b], not specified whether the dose interruption was for a least 7 consecutive days.
Source: ITC report.[56]

B.2.9.4 Discussion

Fedratinib provides superior efficacy benefits compared with BAT, as demonstrated by greater proportions of patients with ≥35% SVR and ≥50% TSS reduction.[56]

Where the efficacy of BAT was informed by the SIMPLIFY-2 study, treatment with fedratinib 400 mg led to a greater proportion of patients achieving ≥35% reduction in SVR (naïve ITC 30.9% vs 5.8%; Table 18) and ≥50% reduction in TSS compared with BAT (naïve ITC [redacted]% vs 5.9%; Table 19).[56] For both endpoints, results were similar when a naïve comparison was performed and when adjustments for prognostic variables were performed. These results indicate the benefit of fedratinib in a population that would otherwise have very poor response rates, even when the proportion of ruxolitinib in the BAT arm is high (89% in SIMPLIFY-2). This supports the feedback from clinicians that ruxolitinib use in this context is suboptimal and does not significantly alter the course of disease.[18]

Where the efficacy of BAT was informed by the PERSIST-2 study, treatment with fedratinib 400 mg led to a greater proportion of patients with a platelet count <100 x 10^9/L achieving ≥35% reduction in SVR ([redacted]% vs 3%) and ≥50% reduction in TSS ([redacted]% vs 15%) compared with BAT (Table 17).[56]

B.2.9.5 Uncertainties in the indirect treatment comparison

B.2.9.5.1 Comparison of fedratinib with BAT, informed by the JAKARTA-2 and SIMPLIFY-2 studies

Both MAIC and STC rely on the strong assumption that all prognostic factors and treatment effect modifiers are required to be known.[57] Identification of treatment effect modifiers was not possible for these analyses given the JAKARTA-2 study is a single-arm trial and there is a paucity of literature on this topic. The variables that could be adjusted for in these analyses were also limited to the reported baseline characteristics from the SIMPLIFY-2 study.

The MAIC methodology, when adjustment is made for ECOG PS, results in an ESS which retains a significant proportion of the original sample size (91.7 compared with the original sample size of 97). However, these analyses are limited by not including adjustment for transfusion dependence, identified by an external haematologist as a baseline characteristic that has a clinically meaningful imbalance between the JAKARTA-2 and SIMPLIFY-2 studies. However, attempts to adjust for transfusion dependence resulted in an ESS of 34.4, therefore estimates using the weights from this adjustment are likely to be unstable. The weights from this adjustment also indicated that a small set of JAKARTA-2 patients were influencing the results.

It should also be noted that the JAKARTA-2 and SIMPLIFY-2 studies used different symptom questionnaires to calculate TSS (JAKARTA-2 uses the modified MF-SAF and SIMPLIFY-2 uses Version 2 of the Myeloproliferative Neoplasm Symptom Assessment Form). Therefore, results from the comparison of the percentages of patients achieving ≥50% reduction in TSS should be interpreted with caution.

B.2.9.5.2 Comparison of fedratinib to BAT informed by the JAKARTA-2 and PERSIST-2 studies

One of the main limitations to the analyses comparing fedratinib with BAT, where the efficacy of BAT is informed by the PERSIST-2 study, is the unavailability of information for the subgroup of BAT-treated PERSIST-2 patients who had received prior ruxolitinib. Information is not available to understand which treatments patients in this subgroup received; therefore, the composition of BAT is unknown. The

baseline characteristics for this subgroup are also not reported meaning it is difficult to conclude how similar patients in JAKARTA-2 and this PERSIST-2 subgroup are. A robust analysis that adjusts for differences in baseline characteristics is also not possible.

The subgroup of JAKARTA-2 patients with a platelet count <100 x 10^9/L was used to compare to the PERSIST-2 evidence. All patients in PERSIST-2 had a platelet count ≤100 x 10^9/L. However, even though the information was not available for the subgroup of PERSIST-2 patients who had received prior ruxolitinib, there is likely to still be a disparity in patients with a platelet count <50 x 10^9/L. JAKARTA-2 only included patients with a platelet count ≥50 x 10^9/L, whereas 44% of the ITT BAT-treated PERSIST-2 patients had a platelet count <50 x 10^9/L.

As with the comparison of SIMPLIFY-2 evidence, the PERSIST-2 study used a different symptom assessment form to that used in JAKARTA-2, meaning results from the comparison of the percentages of patients achieving ≥50% reduction in TSS should be interpreted with caution.

B.2.9.5.3 BAT in comparator studies versus clinical practice

Expert elicitation was sought to establish whether the composition of BAT in PERSIST-2 and SIMPLIFY-2 studies was representative of how patients would be treated with BAT in the UK. Feedback received during the advisory board indicated that PERSIST-2 is not representative of patients receiving ruxolitinib in BAT in the UK as it included patients with platelets <50 x 10^9 /L, for which ruxolitinib is not licensed.[18] Specifically, 34 of the 72 patients in the BAT arm of PERSIST-2 had platelets <50 x 10^9 /L at baseline. Additionally, many patients with lower platelet count must reduce their ruxolitinib dose as per licensing, so the proportion of ruxolitinib in BAT observed in PERSIST-2 (44%) may be more conservative than what would be seen in UK clinical practice.

Considering the above, as well as clinician insights concerning patients rarely being discontinued from ruxolitinib in an attempt to manage prevailing symptoms or mitigate withdrawal symptoms (see Section B.1.3.4), SIMPLIFY-2 was considered a

more realistic representation of ruxolitinib use in BAT (89%) in UK clinical practice.[18] As such, this has been used to inform the economic modelling (see Section B.3.5.1).

B.2.10. Adverse reactions

B.2.10.1 Treatment exposure

The median number of treatment cycles was six (inter quartile range 3.9–8.9).[37] Fourteen (14.4%) patients received more than 12 cycles. Treatment was discontinued due to early study termination in 63 (65%) patients. The remainder of patients discontinued study treatment due to AEs (19%), disease progression (6%), patient decision (3%), or other reasons (7%). Thirty-eight (39%) patients had at least one dose reduction, 13 (13%) had two dose reductions and four (4%) had more than two dose reductions (see Appendix F for further information regarding dose modifications). A total of 25 (25.8%) patients had a dose interruption for at least 7 consecutive days.

Most patients ([redacted]%) received the maximum daily dose of 400 mg fedratinib and almost all patients received ≥ 80% of the intended dose ([redacted]%).[43] A summary of the treatment exposure in JAKARTA-2 is provided in Table 21.

Table 21: Fedratinib exposure (JAKARTA-2, all treated population)

	Fedratinib 400 mg (N=97)
Total number of cycles administered (all patients)	[redacted]
Cycles administered[a]	
Mean (SD)	[redacted]
Median (min, max)	6.0 ([redacted])
Cycles by category, n (%)	
0–3 cycles	[redacted]
> 3–6 cycles	[redacted]
> 6–9 cycles	[redacted]
> 9–12 cycles	[redacted]
> 12 cycles	[redacted]
Duration of exposure[b] (weeks)	
Sum	[redacted]
Mean (SD)	[redacted]
Median (min, max)	[redacted]
Actual dose intensity[c] (mg/week)	

	Fedratinib 400 mg (N=97)
Mean (SD)	[redacted]
Median (min, max)	[redacted]
Maximum dose, n (%)	
400 mg	[redacted]
500 mg	[redacted]
600 mg	[redacted]
800 mg[d]	[redacted]
Cumulative dose, mg	
Mean (SD)	[redacted]
Median (min, max)	[redacted]
Number (%) of patients with ≥80% intended dose[e]	[redacted]

Key: max, maximum; min, minimum; SD, standard deviation.
Notes: [a], a cycle was counted if the patient received at ≥ 1 (even partial) dose of fedratinib; [b], duration of exposure in weeks was defined as: (last dose date − first dose date + 1)/7; [c], the actual dose intensity was defined as the cumulative dose divided by duration of fedratinib exposure in terms of the number of weeks on study; [d], one patient took an 800 mg total daily dose on C4D8 instead of 400 mg. This was reported as an accidental overdose. Fedratinib was interrupted for 1 day. The overdose was also captured as an adverse event of special interest; [e], treatment compliance was defined as the total actual dose taken divided by total intended dose (reported number of days on treatment excluding interruptions).
Source: JAKARTA-2 CSR[43] and Harrison et al 2017.[37]

B.2.10.2 Summary safety data

The safety analyses were performed in the all treated population; defined as enrolled patients who took at least one dose (even if partial) of study medication (n=97).[43]

All 97 patients had at least one treatment-emergent adverse event (TEAE) of any grade.[43, 46] Grade 3 or 4 TEAEs were reported by [redacted] (63%) patients, including transfusion dependency in [redacted] ([redacted]%) patients. Treatment-emergent serious adverse events (SAEs) were reported by [redacted] (34%) patients. Seven (7%) patients had a TEAE that led to death during treatment or follow-up; in four cases, the cause of death was determined to be due to disease progression and the other three cases were due to a TEAE considered not related to study treatment. TEAEs leading to treatment discontinuation occurred in [redacted] (20%) patients and TEAEs leading to dose modification occurred in [redacted] ([redacted]%) patients (see Appendix F for further information regarding TEAEs leading to dose modifications).

An overview of the TEAEs associated with fedratinib in JAKARTA-2 is provided in Table 22.

Table 22: Safety overview (JAKARTA-2, all treated population)

n (%)	Fedratinib 400 mg (N = 97)
TEAE	97 (100%)
Treatment-related TEAE	
Grade 3 or 4 TEAEs	61 (63%)
Treatment-related Grade 3 or 4 TEAEs	
TEAE leading to death	7 (7%)
Treatment-related TEAE leading to death	0
Treatment-emergent SAEs	33 (34%)
Treatment-related treatment-emergent SAEs	
TEAEs leading to permanent treatment discontinuation	19 (20%)
TEAEs leading to dose modification	

Key: SAE, serious adverse event; TEAE, treatment-emergent adverse event.
Notes: Data are for patients with ≥ 1 TEAE.
Source: Harrison 2020[46], and JAKARTA-2 CSR.[43]

B.2.10.3 Common adverse event data

The most common non-haematological TEAEs were gastrointestinal disorders including diarrhoea in 60 (62%) patients, nausea in 54 (56%) patients, vomiting in 40 (41%) patients, constipation in 20 (21%) patients and abdominal pain in 12 (12%) patients.[37] Other common non-haematological TEAEs in other system order classes included pruritus in 17 (17.5%) patients, fatigue in 15 (15.5%) patients, cough and headache in 13 (13%) patients each, urinary tract infection and dyspnoea in 12 (12%) patients each and dizziness in 11 (11%) patients.

The most common haematological TEAEs were anaemia in 47 (48%) patients and thrombocytopenia in 26 patients (27%).[37] Grade 3 or 4 anaemia was reported in 37 (38%) patients and thrombocytopenia in 21 (22%) patients.

A summary of the common AEs reported in JAKARTA-2 is presented in Table 23. For details of patients with ≥ 1 TEAE of any grade, see Appendix F.

Table 23: Common adverse events (JAKARTA-2, all treated population)

	Fedratinib 400 mg (N=97)	
	Grade 1–2	Grade 3–4
Haematological adverse events, n (%)		
Anaemia	10 (10%)	37 (38%)

	Fedratinib 400 mg (N=97)	
	Grade 1–2	**Grade 3–4**
Thrombocytopenia	5 (5%)	21 (22%)
Lymphopenia	1 (1%)	3 (3%)
Non-haematological adverse events, n (%)		
Diarrhoea	56 (58%)	4 (4%)
Nausea	54 (56%)	0
Vomiting	40 (41%)	0
Constipation	19 (20%)	1 (1%)
Pruritus	16 (16%)	0
Fatigue	13 (13%)	2 (2%)
Headache	12 (12%)	1 (1%)
Cough	13 (13%)	0
Urinary tract infection	12 (12%)	0
Dyspnoea	11 (11%)	1 (1%)
Dizziness	11 (11%)	0
Abdominal pain	7 (7%)	2 (2%)
Alanine aminotransferase increased	3 (3%)	3 (3%)
Pneumonia	3 (3%)	2 (2%)
Hyperlipasaemia	1 (1%)	3 (3%)
Hyperuricaemia	2 (2%)	2 (2%)
Dehydration	1 (1%)	2 (2%)
Tumour lysis syndrome	0	2 (2%)
Cardiac failure	1 (1%)	2 (2%)
Amylase increased	1 (1%)	2 (2%)
Blood bilirubin increased	0	2 (2%)
Cardiac failure	1 (1%)	2 (2%)
Respiratory failure	0	0
Splenic rupture	0	0

Notes: Shown are any grade event occurring in more than 10% of patients and Grade 3–4 events occurring in more than one patient.
Source: Harrison et al 2017[37], and Harrison 2020.[46]

Common AE data for the Stringent Criteria and Sensitivity Cohorts is presented in Appendix D.

B.2.10.4 Treatment-emergent SAEs

Treatment-emergent SAEs were reported in 33 (34%) patients.[37, 43] The most common SAE was cardiac disorders, reported in five patients (5%). Pneumonia was

reported in four patients (4%), pleural effusion in three (3%) and fall in [redacted] patients ([redacted]).

[redacted] patients ([redacted]%) had SAEs considered treatment related.[43] Pneumonia was the only treatment-related SAE reported in more than one patient and occurred in [redacted] patients.

B.2.10.5 Adverse events leading to treatment discontinuation

TEAEs leading to treatment discontinuation occurred in [redacted] (20%) patients, of whom [redacted] ([redacted]%) had a Grade 3 or 4 event.[43, 46] The most common reason for treatment discontinuation was Grade 3 or 4 thrombocytopenia, which occurred in two patients. One patient had disease transformation to AML, which was considered an AE, but the reason for discontinuation was recorded as disease progression.

One case of Grade 3 encephalopathy was reported, it was subsequently determined by an independent expert safety panel to be related to hepatic encephalopathy and inconsistent with WE.[37] The event resolved within one week after discontinuation of fedratinib treatment.

A summary of treatment-emergent and treatment related adverse events leading to permanent treatment discontinuation is provided in Table 24.

Table 24: Treatment-emergent and treatment-related AEs leading to treatment discontinuation (JAKARTA-2; all treated population)

System organ class, preferred term, n (%)[a]	Fedratinib 400 mg (N = 97)			
	All grades	Grade 3 or 4	Treatment-related	
			All grades	Grade 3 or 4
Patients with ≥ 1 TEAE leading to permanent treatment discontinuation	[redacted]	[redacted]	10 (10%)	8 (8%)
Gastrointestinal disorders	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	2 (2%)	2 (2%)
Abdominal discomfort	[redacted]	[redacted]	[redacted]	[redacted]
Nausea	[redacted]	[redacted]	1 (1%)	0
Vomiting	[redacted]	[redacted]	1 (1%)	0
Investigations	[redacted]	[redacted]	[redacted]	[redacted]
Blood creatinine increased	[redacted]	[redacted]	1 (1%)	0

System organ class, preferred term, n (%)[a]	Fedratinib 400 mg (N = 97)			
	All grades	Grade 3 or 4	Treatment-related	
			All grades	Grade 3 or 4
Gamma-glutamyltransferase increased			1 (1%)	1 (1%)
Platelet count decreased			1 (1%)	1 (1%)
Weight decreased			1 (1%)	1 (1%)
Blood and lymphatic system disorders				
Thrombocytopenia			1 (1%)	1 (1%)
Anaemia			1 (1%)	1 (1%)
Cytopenia				
Febrile neutropenia				
Thrombotic thrombocytopenic purpura			1 (1%)	1 (1%)
Cardiac disorders				
Atrial fibrillation				
Cardio-respiratory arrest				
Infections and infestations				
Pneumonia				
Sepsis				
General disorders and administration site conditions				
Fatigue				
Injury, poisoning and procedural complications				
Splenic rupture				
Metabolism and nutrition disorders				
Hyperlipasaemia				
Neoplasms benign, malignant and unspecified (including cysts and polyps)				
Transformation to AML				
Nervous system disorders				
Encephalopathy				
Vascular disorders				
Shock				

Key: AML, acute myeloid leukaemia; MedDRA, Medical Dictionary for Regulatory Activities; NCI-CTCAE, National Cancer Institute Common Terminology Criteria for Adverse Events; PT, preferred term; SOC, system order class; TEAE, treatment-emergent adverse event.
Notes: [a], SOCs and PTs were coded using the MedDRA Version 20.1. If multiple TEAEs were reported within a given PT, only one event was counted per patient. Table sorted by decreasing frequency of SOC and PT in all grades column of TEAEs (without consideration of relatedness).
Source: JAKARTA-2 CSR[43] and Harrison 2020.[46]

See Appendix F for a summary of the AEs leading to dose reduction or interruption in JAKARTA-2.

B.2.10.6 Adverse events leading to death

Seven (7%) patients died during treatment in JAKARTA-2, but none of the deaths was deemed to be related to fedratinib.[46] Three patients died due to fatal TEAEs of pneumonia, shock and cardiorespiratory arrest. The four other patients died due to disease progression as the main cause of death.

A summary of TEAEs leading to death is provided in Table 25.

Table 25: TEAEs leading to death (JAKARTA-2, all treated population)

System organ class preferred term, n (%)[a]	Fedratinib 400 mg (N = 97)
Patients with at least one TEAE leading to death	7 (7%)
General disorders and administration site conditions	
Disease progression[b]	
General physical health deterioration	
Infections and infestations	
Pneumonia	
Sepsis	
Cardiac disorders	
Cardio-respiratory arrest	
Neoplasms; benign, malignant, and unspecified (including cysts and polyps)	
Acute myeloid leukaemia	
Respiratory, thoracic, and mediastinal disorders	
Respiratory failure[b]	
Vascular disorders	
Shock	

Key: MedDRA, Medical Dictionary for Regulatory Activities; PT, preferred term; SOC, system order class; TEAE, treatment-emergent adverse event.
Notes: [a], SOCs and PTs were coded using the MedDRA Version 20.1. If multiple TEAEs were reported within a given PT, only one event was counted per patient; [b] one patient had two TEAEs leading to death; disease progression and respiratory failure. TEAEs leading to death were those that occurred during the on-treatment period (the time from first dose of fedratinib to 30 days after last dose of fedratinib). The table is sorted by decreasing frequency of SOC and PT.
Source: Harrison 2020,[46] and JAKARTA-2 CSR.[43]

B.2.10.7 Safety overview

The most common TEAEs observed in JAKARTA-2 were consistent with the known safety profile of fedratinib, could be managed with dose modifications and were not a frequent reason for discontinuation of fedratinib.

The most frequent Grade 3 or 4 events in this study were anaemia and thrombocytopenia.[43] Given that the patients in the study tended to have advanced disease, were heavily pre-treated and had higher rates of baseline anaemia and thrombocytopenia, this finding is not unexpected. Additionally, as the JAK/STAT pathway modulates haematopoiesis it may potentially be a contributing factor to cytopenias. The three fatal TEAEs (pneumonia, cardio-respiratory arrest and shock) were not considered to be related to fedratinib treatment.[46]

Analysis of the signs and symptoms that may be associated with events of WE in JAKARTA-2 were not suggestive of any confirmed cases. Increased clinical awareness of the potential for developing WE and routine thiamine monitoring, with thiamine replacement as appropriate, sufficiently minimises the risk of developing this AE.

B.2.11. Ongoing studies

The Phase III, single-arm FREEDOM study of fedratinib in patients with DIPSS intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib is currently recruiting and due to read out in 2022 (US study).

The Phase III FREEDOM-2 study of fedratinib compared with BAT in patients with DIPSS intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib is currently recruiting and due to read out in 2022.

B.2.12. Innovation

There is a notable absence of recent innovation in myelofibrosis, with no new therapies approved in Europe since ruxolitinib in 2012.

Ruxolitinib is currently the only targeted therapy for patients with myelofibrosis that has been appraised by NICE; however, more than 50% do not maintain responses or are intolerant to ruxolitinib long-term.[38] In patients discontinuing ruxolitinib and receiving BAT, the estimated median OS ranges between 13–16 months.[7, 15-17] Those who continue suboptimal ruxolitinib are expected to have a similar survival as those observed in the literature, which is supported by the clinical experts.[18]

Despite the poor prognosis and high symptom burden in patients who have been treated with ruxolitinib, the standard of care is currently limited to BAT, which is not associated with significant SVR or TSS reduction.[60] This demonstrates a clear unmet need for a step change in the current myelofibrosis treatment pathway – so that a safe and efficacious, targeted therapy can be offered to patients who otherwise have poor outcomes.

Fedratinib selectively inhibits JAK2, with higher inhibitory activity for JAK2 over family members JAK1, JAK3 and TYK2 and is a more selective inhibitor of JAK2 than ruxolitinib. It is the only JAK inhibitor with demonstrable efficacy in a population that are relapsed, refractory, or intolerant to ruxolitinib who have a high unmet need. JAKARTA-2 has shown that treatment with fedratinib is associated with a substantial and clinically significant SVR and TSS reduction in patients treated with ruxolitinib.[37, 43, 44] This efficacy is further supported by similar results in a ruxolitinib naïve population (JAKARTA).[48] This is an unprecedented finding given the considerably worse prognosis for ruxolitinib treated versus ruxolitinib naïve patients.

Fedratinib offers relief of debilitating symptoms associated with myelofibrosis, as reflected by improvements in HRQoL measures of physical functioning, fatigue, pain, insomnia and appetite loss.[37, 43, 44] These improvements alleviate the burden of disease experienced by both patients and their loved ones, and better enables patients to carry out their normal daily functions.

Fedratinib offers an alternative, convenient, well tolerated oral therapy that delivers clinically meaningful outcomes, including a survival gain for patients treated with ruxolitinib. Fedratinib, therefore, offers a step change in the clinical treatment pathway for myelofibrosis patients.

B.2.13. Interpretation of clinical effectiveness and safety evidence

JAKARTA-2 demonstrated that treatment with fedratinib is associated with considerable reductions in spleen volume and size, as well as marked improvements to symptoms in patients previously treated with ruxolitinib.

Splenomegaly is the key physical feature and cause of symptoms of myelofibrosis, as such SVR forms an important treatment goal. Internationally recognised research groups have identified ≥ 35% SVR as the appropriate threshold for defining response in patients with myelofibrosis[51], about a third (31%) of patients in JAKARTA-2 achieved this response.[44]

In lieu of availability of curative treatments, the relief of debilitating symptoms is another important treatment goal in myelofibrosis. The clinically meaningful threshold for symptom response is ≥50% reduction in TSS,[51] with [redacted]% of patients in JAKARTA-2 having achieved this.[44] Additionally, fedratinib was associated with considerable median percent changes of the key symptoms compromising the TSS; including pain under ribs ([redacted]%), night sweats ([redacted]%), early satiety ([redacted]%), abdominal discomfort ([redacted]%), pruritis ([redacted]%) and bone or muscle pain ([redacted]%).[43] Alleviating these symptoms provides patients with an improved ability to carry out normal daily functions and relieves some of the physical and psychological burden associated with myelofibrosis. Indeed, the impact of fedratinib on debilitating symptoms is further supported by improvements in HRQoL, with half of evaluable patients having demonstrated postbaseline improvements in global QoL, physical functioning, fatigue, pain, insomnia and appetite loss.[43]

Although there is a paucity of direct comparative evidence in the population treated with ruxolitinib, a superior comparative effect for fedratinib versus placebo is provided by JAKARTA. In JAKARTA, fedratinib 400 mg was associated with ≥ 35% SVR in 36% of patients compared with 1% of patients receiving placebo; and ≥ 50% reduction in TSS in 36% of patients compared with 7% of patients receiving placebo.[48]

Comparative data for fedratinib versus standard of care is informed by ITC and supports the additional benefit of fedratinib versus BAT in terms of ≥ 35% SVR and

≥ 50% TSS reduction.[56] When the efficacy of BAT was informed by the PERSIST-2 or SIMPLIFY-2 studies, treatment with fedratinib 400 mg led to greater proportions of patients achieving ≥ 35% reduction in SVR and ≥ 50% reduction in TSS compared with BAT.

Owing to the early termination of JAKARTA-2 following the clinical hold, which mandated cessation of fedratinib therapy, survival data in JAKARTA-2 and JAKARTA were immature; however, survival was assessed as an exploratory outcome. Both trials reported similar OS rates for fedratinib 400 mg at 12 months ([redacted]% for JAKARTA-2 and 91.6% for JAKARTA).[49] In JAKARTA, the HR for fedratinib 400 mg versus placebo was [redacted] (95% CI [redacted], [redacted]), indicating a trend for prolonging OS.[49] There is clinical plausibility to suggest that this trend may also apply to the population that have been treated with ruxolitinib, given the similarity in OS results (and other efficacy outcomes) between JAKARTA-2 and JAKARTA studies.

In the evaluation of the clinical and economic effectiveness of fedratinib, the inclusion of intermediate-1 patients with symptoms from JAKARTA-2 reflects a slightly broader patient population than those anticipated to receive fedratinib in UK clinical practice; however, it enables the best fit for comparison against other studies in the ITC (see Section B.2.9). The post-hoc analyses of Int-2/high-risk patients from JAKARTA-2 demonstrated efficacy results that were consistent with the ITT population.

Crucially, the inclusion of intermediate-1 patients provides the most conservative estimate of the efficacy of fedratinib and therefore use of the ITT population in the ITC and economic modelling are not thought to bias in favour of fedratinib. This indicates that the treatment benefit of fedratinib in UK patients may be greater than that demonstrated in the clinical and economic evaluations.

The proposed position of fedratinib in the treatment pathway is narrower than the marketing authorisation because the population of patients who have been treated with ruxolitinib represents the greatest unmet need in myelofibrosis, for which the clinical and cost-effectiveness of fedratinib is most demonstrable. The survival

outcomes in patients who have been treated with ruxolitinib are poor, with studies indicating a median OS of 13–16 months following ruxolitinib discontinuation.[7, 15-17]

This highlights the need for a treatment such as fedratinib to not only alleviate the debilitating symptoms associated with myelofibrosis, improve HRQoL but to offer an opportunity for life extension in a disease state that fulfils end of life criteria (Table 26).

Table 26: End-of-life criteria

Criterion	Data available	Reference in submission (section and page number)
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	Several published reports for patients that have been treated with ruxolitinib demonstrate a median OS of 13–16 months.[7, 15-17] These estimates are likely to be even lower in the intermediate-2 and high-risk population.	Section B.3.3, p 119
There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment	The economic model predicts 0.85 additional life years (approximately 10 months) with fedratinib compared to best available therapy. The OS effect of fedratinib can be supported by findings from JAKARTA, which indicate that compared with placebo, treatment with fedratinib improves OS.[49] In JAKARTA, treatment with fedratinib 400 mg was associated with an OS HR of [redacted] (95% CI: [redacted]; p = [redacted]) versus placebo	Section B.2.6, p 48
Key: CI, confidence interval; HMRN, Haematological Malignancy Research Network; HR, hazard ratio; NHS, National Health Service; OS, overall survival.		

B.3. Cost effectiveness

B.3.1. Published cost-effectiveness studies

An SLR was performed to identify published cost-effectiveness studies in myelofibrosis to support the development of a de novo economic model for fedratinib. The search strategy and study selection criteria are described in detail in Appendix G.

In total, 1,126 potentially relevant papers were identified in database searches. After exclusion of irrelevant articles (n = 1,120), and the addition of relevant articles from bibliographic (n = 1) and health technology assessment (HTA) (n = 8) searches, a total of 15 publications were included. As some studies were associated with multiple publications, secondary publications were combined; this resulted in inclusion of nine studies identified from 15 publications.[7, 61-68] Table 27 and Table 28 present a summary of the nine cost-effectiveness studies identified by the SLR.

Table 27: Characteristics of published cost-effectiveness studies

Study name	Country	Type of study, Type of model	Cost year, Currency, Discount rate	Health economic perspective, Time horizon, Cycle length	Model health states and definition
Rojas et al., 2016 [61]	Chile	• CUA • Markov cohort state transition model	• NR • US dollar (US$) • 3%	• Chilean public healthcare system • Lifetime • NR	NR
Vandewalle et al., 2016 [62]	Portugal	• CEA • Discrete state cohort model	• 2013 • Euro (€) • 5%	• Portuguese National Health Service • Lifetime • 4 weeks	NR
Hahl et al., 2015 [63]	Finland	• CUA • Survival-based decision analytic cohort model	• NR • Euro (€) • 3%	• Finnish healthcare payer perspective • Lifetime • NR	• Alive: on-treatment • Alive: off-treatment • Death
NICE [Ruxolitinib], 2016 [7]	UK	• CUA • Individual patient discrete event simulation model (de novo model)	• 2015 • Pound (£) • 3.5%	• NHS and Personal Social Services perspective • Lifetime • Not applicable although reported as 1 week	Four health states: • Alive on ruxolitinib • Alive on BAT • Alive on supportive care • Death
PBAC [Ruxolitinib], 2015 [64]	Australia	• CUA • NR	• NR • Australian dollar (A$) • NR	• NR • 20-year • 24 weeks	• Baseline state with controlled pain/fatigue

Study name	Country	Type of study, Type of model	Cost year, Currency, Discount rate	Health economic perspective, Time horizon, Cycle length	Model health states and definition
					• Spleen response with controlled pain/fatigue • No spleen response but controlled pain/fatigue • Death
SMC [Ruxolitinib], 2015 [65]	Scotland	• CUA • Discrete event simulation model	• NR • Pound (£) • NR	• NHS perspective • Lifetime • NA	• Ruxolitinib • BAT • Supportive care • Palliative care • Death
CADTH [Ruxolitinib], 2013 [66]	Canada	• CUA • NR	• NR • Canadian dollar (C$) • NR	• Publicly funded healthcare system • 96- to 144-week (Economic Guidance Panel reanalyses assumed the model's time horizon to be shorter than the proposed lifetime time horizon) • NR	• NR

Study name	Country	Type of study, Type of model	Cost year, Currency, Discount rate	Health economic perspective, Time horizon, Cycle length	Model health states and definition
NCPE [Ruxolitinib], 2013 [67]	Ireland	• CUA • Markov cohort state transition model	• NR • Euro (€) • NR	• NR • 35 years • 12 weeks	Four health states: • Responder • Non-responder • Discontinuation • Death
El Ouagari et al., 2012 [68]	Canada	• CUA • Markov model	• NR* • Canadian dollar (C$) • NR	• Canadian Societal Perspective • Lifetime • 12 weeks	Four health states: • Responder • Non-responder • Leukaemic transformation • Death
Key: BAT, best available therapy; CEA, cost-effectiveness analysis; CUA, cost–utility analysis; NHS; National Health Service; NR, not reported.					

Table 28: Results of published cost-effectiveness studies

Study	Treatments	Sources used to measure effectiveness	Outcomes	Costs	ICERs
Rojas et al., 2016 [61]	• Ruxolitinib • BAT	NR	Incremental QALYs: 0.98	Total costs • Ruxolitinib: US$101,926 • BAT: US$47,070	ICER: US$54,500/QALY
Vandewalle et al., 2016 [62]	• Ruxolitinib • BAT	COMFORT-II[69]	Total LYG: Discounted at a 5% annual rate • Ruxolitinib: 5.39 • BAT: 2.96	Total costs: • Ruxolitinib: €188,967 • BAT: €91,915	ICER: €40,000/LY
Hahl et al., 2015 [63]	• Ruxolitinib • BAT	COMFORT-II[69]	Incremental QALYs: 2.43	Incremental costs: €102,802	ICER: €42,367/QALY
NICE [Ruxolitinib], 2016 [7]	• Ruxolitinib • BAT	COMFORT-I[70] and COMFORT-II[69]	LYG • BAT: 2.15 • Ruxolitinib: 5.96 QALYs • BAT: 1.476 • Ruxolitinib: 3.989	Total costs (with PAS) • BAT: £36,271 • Ruxolitinib: £149,114	Base-case results (with PAS) ICER: £44,905/QALY

Study	Treatments	Sources used to measure effectiveness	Outcomes	Costs	ICERs
PBAC [Ruxolitinib], 2015 [64]	• Ruxolitinib • Placebo	COMFORT-I[70] and COMFORT-II[69]	Results of the economic evaluation LYG (Intermediate-1 patients) • Ruxolitinib: 5.015 • Placebo: 2.389 QALYs (Intermediate-1 patients) • Ruxolitinib: 3.163 • Placebo: 0.936 LYG (Intermediate-2 or high-risk patients) • Ruxolitinib: 5.015 • Placebo: 2.389 QALYs (Intermediate-2 or high-risk patients) • Ruxolitinib: 3.163 • Placebo: 0.936	Cost for QALY gained Intermediate-1 patients: • Ruxolitinib: redacted • Placebo: A$10,356 Intermediate-2 or high-risk patients • Ruxolitinib: redacted • Placebo: A$10,822	ICER for intermediate-2/high risk patients: A$45,000 to A$75,00
SMC [Ruxolitinib], 2015 [65]	• Ruxolitinib • BAT	COMFORT-I[70] and COMFORT-II[69]	Base-case QALYs results Incremental: 1.99	Base case costs results (with PAS) Incremental costs: £98,982	Base-case ICER per QALYs results (with PAS): £49,774
CADTH [Ruxolitinib], 2013 [66]	• Ruxolitinib • BAT	NCT00509899[71]	Incremental QALYs: 0.06 to 0.07	Incremental costs: C$14,634 to C$14,679	ICER: C$199,118 to C$259,698

Study	Treatments	Sources used to measure effectiveness	Outcomes	Costs	ICERs
NCPE [Ruxolitinib], 2013 [67]	• Ruxolitinib • BAT •	COMFORT-I[70] and COMFORT-II[69]	Incremental QALYs: 1.20	Incremental costs: €84,292	ICER: €70,252
El Ouagari et al., 2012 [68]	• Ruxolitinib • BAT	COMFORT-II[69] and NCT00509899[71]	QALYs: • Ruxolitinib: 4.01 • BAT: 2.82	Total costs: • Ruxolitinib: C$494,859 • BAT: C$421,755	ICER: C$61,444

Key: BAT, best available therapy; CADTH, Canadian Agency for Drugs and Technologies in Health; ICER, incremental cost-effectiveness analysis; LYG, life-years gained; NCPE, National Centre for Pharmacoeconomics; NICE, National Institute for Health and Care Excellence; NR, not reported; PAS, patient access scheme; PBAC, Pharmaceutical Benefits Advisory Committee; QALY, quality-adjusted life year; SMC, Scottish Medicines Consortium.

B.3.2. Economic analysis

The SLR included nine studies from 15 publications which investigated the cost-effectiveness of therapies in patients with myelofibrosis. All nine studies assessed the cost-effectiveness of ruxolitinib relative to either BAT or placebo. The studies included five ruxolitinib HTA submission documents:

- Canada – Canadian Agency for Drugs and Technologies in Health (CADTH), 2013 [66]
- Ireland – National Centre for Pharmacoeconomics (NCPE), 2013 [67]
- England and Wales – National Institute for Health and Care Excellence (NICE), 2016 [7]
- Scotland – Scottish Medicines Consortium (SMC), 2015 [65]
- Australia – Pharmaceutical Benefits Advisory Committee (PBAC), 2015.[64]

The remaining four studies were published in peer-reviewed journals.[61-63, 68]

Where reported, alive health states were often defined by the treatment received (n = 3).[7, 63, 65] This tended to consist of ruxolitinib, BAT, or supportive care. Alternatively, response or non-response were used to define health states (n = 3).[64, 67, 68] One study included leukaemic transformation as a health state (n = 1).[68] The omission of acute myeloid leukaemia as a distinct health state was queried by the evidence review group in NICE TA386 (ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis).[7] Three studies did not explicitly report health states (n = 3).[61, 62, 66]

Where reported, effectiveness outcomes were informed by one or more of three studies: COMFORT-I[70], COMFORT-II[69], and NCT00509899.[71] The primary outcomes of COMFORT-I and COMFORT-II were ≥ 35% reduction in spleen volume from baseline at 24 weeks and 48 weeks, respectively. COMFORT-I also investigated symptom response, as assessed by the TSS of the modified MF-SAF v2.0. NCT00509899 measured the proportion of patients with ≥ 35% reduction in spleen volume from baseline at time intervals up to 48 weeks, and the change in total symptom score from baseline at 24 weeks. In the economic models, six studies used one or both of the COMFORT-I and COMFORT-II studies to inform treatment

response.[7, 62-65, 67] One study used both COMFORT-II and NCT00509899, but did not leverage the symptom score data.[68] Another study used NCT00509899 trial data alone and considered both spleen volume and symptom response to produce their economic recommendations.[66] One study did not report the data source used.[61]

Cohort models were commonly applied (n = 5).[61-63, 67, 68] Patient-level discrete event simulation (DES) was also leveraged (n = 2).[7, 65] Two studies did not report the model type (n = 2).[64, 66]

One previous submission to NICE in myelofibrosis was identified, TA386.[7] The modelling approach for NICE TA386 was adapted for this submission to evaluate the cost-effectiveness of fedratinib in patients with myelofibrosis who have been treated with ruxolitinib. Health states in the model for this submission are defined by treatment, with outcomes and transitions between health states driven by response.

B.3.2.1 Patient population

The main population in the economic analysis comprises adults with disease-related splenomegaly caused by primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis, who have been treated with ruxolitinib and are classified as intermediate-2 or high risk by DIPSS.

The JAKARTA-2 trial is the primary source of evidence for fedratinib outcomes in patients with prior ruxolitinib treatment (N = 97). The ITT population of JAKARTA-2 included patients classified as ‘intermediate-1 with symptoms’ (n = 16, N = 97, ~17%). In the UK, ruxolitinib is recommended and reimbursed only for patients with intermediate-2 and high-risk disease.[7] Given that fedratinib is positioned in those who have been treated with ruxolitinib, the intermediate-2 and high-risk patients in JAKARTA-2 inform the majority of base case inputs of the economic analysis (N = 81). The full JAKARTA-2 ITT population was included in the economic model for use within scenarios and within statistical analyses where the base case population could not reasonably be used due to sample size restrictions.

The primary endpoint of the JAKARTA-2 trial was to evaluate the efficacy of fedratinib based on the reduction of spleen volume at EOC6. The original analysis applied the LOCF method to impute missing EOC6 data. Additionally, outcomes

were initially reported using the PP population (N = 83), which only included patients with baseline and ≥1 post-baseline MRI/CT scan of spleen volume. However, due to the statistical bias associated with the LOCF method, all output was updated to provide conservative and plausible estimates of efficacy in the ITT population without LOCF adjustment.[46] No LOCF is applied within this submission.

In addition, a post-hoc analysis of JAKARTA-2 ITT has been considered for patients who are relapsed, refractory or intolerant to ruxolitinib as per updated definitions in the JAKARTA-2 CSR addendum (termed the Stringent cohort).[72] The definitions of this cohort are described below. For reference, response to ruxolitinib is defined as: ≥ 50% reduction in spleen size for baseline spleen > 10 cm (or ≥ 35% reduction in spleen volume from baseline); non-palpable spleen for baseline spleen between 5 and 10 cm; not eligible for spleen response for baseline spleen < 5 cm

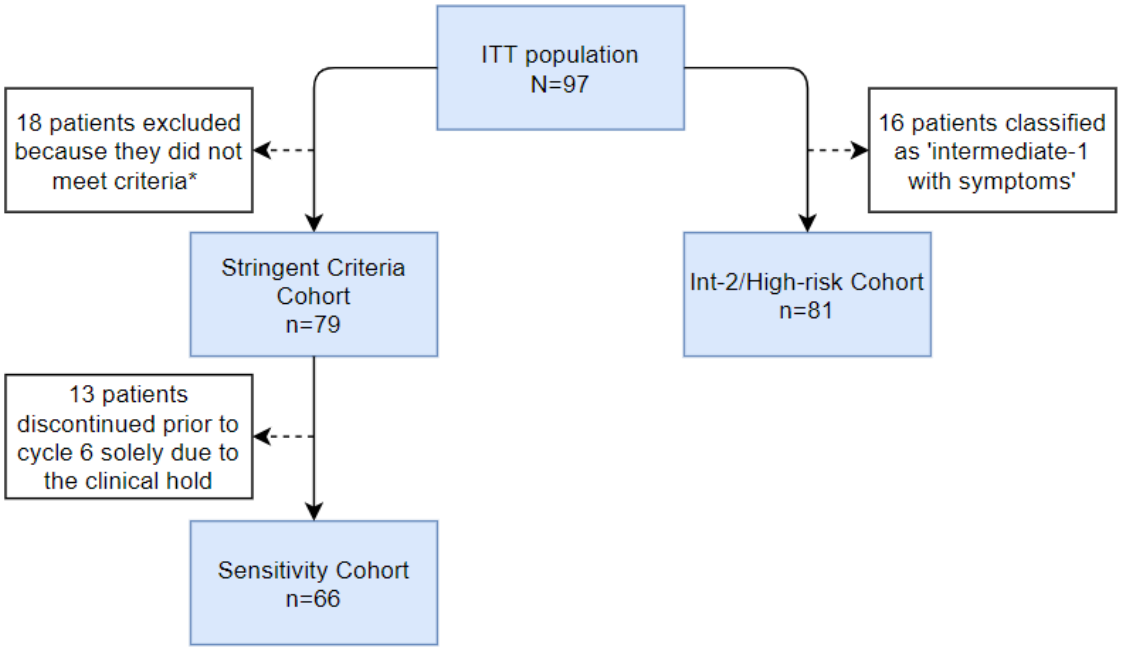
- Relapsed:
  - < 30% reduction in spleen size (or < 10% reduction in spleen volume) at the end of ruxolitinib treatment compared to baseline after an initial response (as defined above). Patients must have had treatment with ruxolitinib for ≥ 3 months
- Refractory:
  - < 30% reduction in spleen size (or < 10% reduction in spleen volume) at the end of ruxolitinib treatment compared to baseline and failure to meet criteria for response (as defined above) during ruxolitinib treatment. Patients must have had treatment with ruxolitinib for ≥ 3 months
- Intolerance:
  - Ruxolitinib treatment for ≥ 28 days complicated by either (i) the development of RBC transfusion requirement (≥ 2 units/month for 2 months), or (ii) toxicity defined as Grade ≥ 3 adverse events (AEs) of thrombocytopenia, anaemia, haematoma, and/or haemorrhage while on treatment with ruxolitinib

In 2013, the Food and Drug Administration placed fedratinib under a clinical hold due to reported cases of WE. Consequently, clinical trials investigating fedratinib, such as JAKARTA-2, were terminated prematurely and patients were mandated to come off treatment. To assess the potential impact of the clinical hold, the Stringent cohort

was further analysed to include only those patients who were not forced to discontinue fedratinib prematurely (termed the Sensitivity cohort). The treatment response observed in the Sensitivity cohort is applied in scenario analysis in the economic model.

Figure 13 provides a diagram of the ITT population and its sub-populations.

Figure 13: JAKARTA-2 intention-to-treat and sub-populations



Key: N, total number of patients; n, number of patients in sub-population
Notes: *Stringent Cohort criteria are relating to the relapsed, refractory and intolerance criteria described in the JAKARTA-2 CSR addendum[72] and above.

B.3.2.2 Model structure

B.3.2.2.1 Model type

The cost-effectiveness analysis for fedratinib is a DES model built in Microsoft Excel®. The modelling methodology emulates the approach taken in the previous technology appraisal of ruxolitinib (TA386), the first JAK-inhibitor approved for myelofibrosis.

In a DES model, patient pathways can be estimated for individuals by sampling directly from time-to-event curves. Therefore, a DES approach does not impose

assumptions that force events to only occur at defined intervals known as 'time-cycles', which are the norm in cohort-based models and many patient-level simulations. The DES approach allows greater flexibility in when transitions can occur and in how transitions are calculated. The model type also enables 'memory' to be implemented – meaning a patient's experience of a health state is recorded so it can more easily influence costs, utilities, and transitions to future health states.

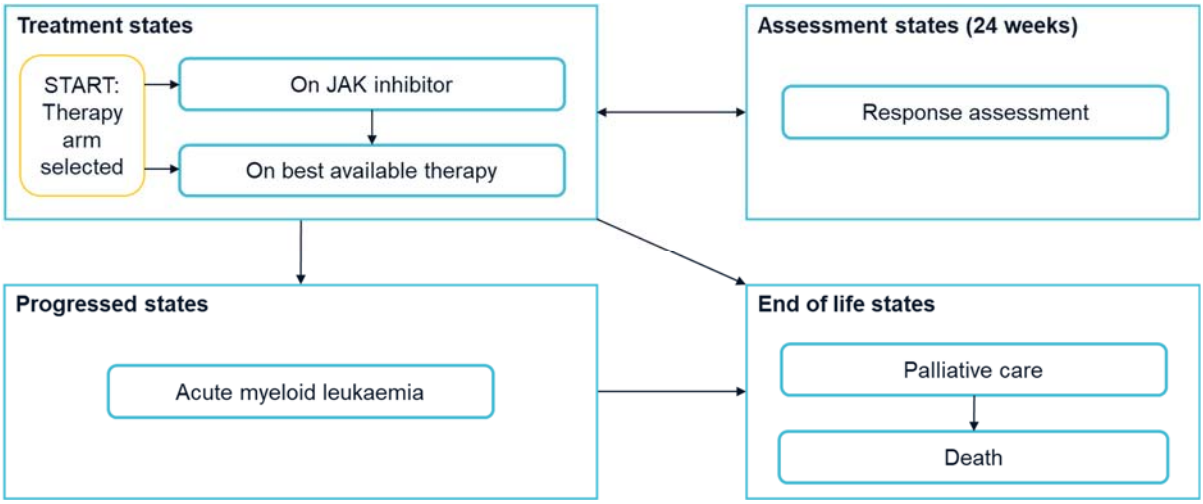
Due to the similarities between the model built for this submission and the model built for TA386, previously stated advantages of the DES approach in myelofibrosis are also applicable to this decision problem. These include its flexibility in handling:

- A response assessment at 24 weeks
- The progressive nature of disease and transition to subsequent treatments
- Worsening health-related quality of life within a health-state (if applicable)
- Ease of adaptation to explore alternative structural assumptions

B.3.2.2.2 Model structure

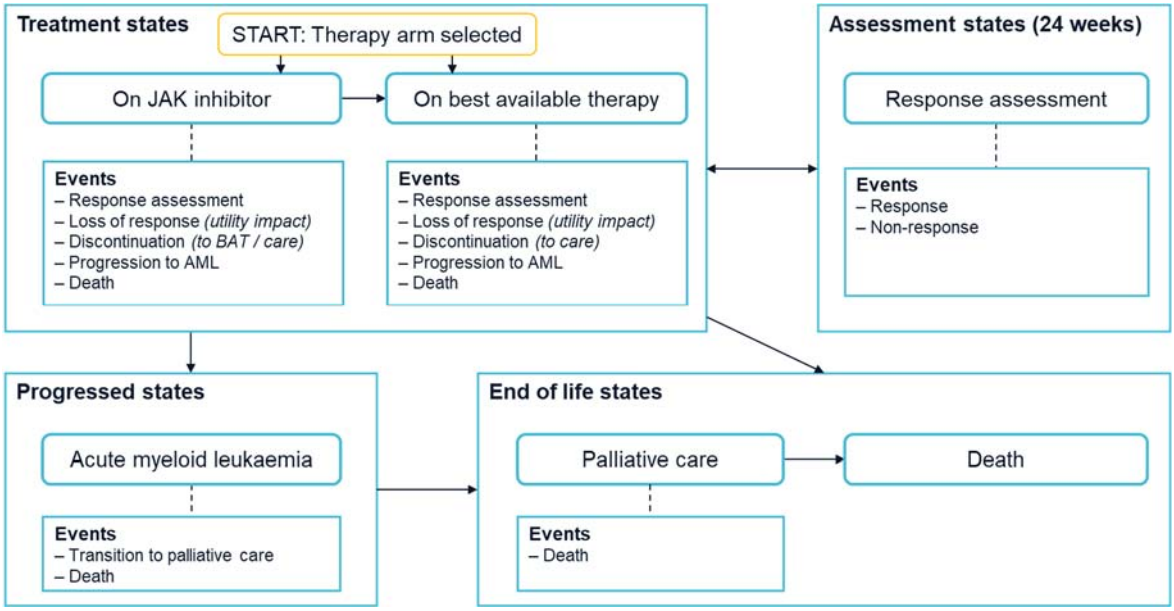
A simplified model diagram is provided in Figure 14. The simplified diagram displays the health states and transitions in the model. A more detailed diagram is provided in Figure 15, which includes the potential events for each health state.

Figure 14: Simplified model diagram



Key: JAK, Janus kinase.

Figure 15: Detailed model diagram with events



Key: AML, acute myeloid leukaemia; BAT, best available therapy; JAK, Janus kinase.

The health states in the model are broken down into four categories: treatment states, assessment states, progressed states, and end-of-life states.

In treatment states, patients are either on fedratinib or BAT. Patients accrue costs according to the treatment received and accrue quality-adjusted life years (QALYs) in line with their response to treatment.

The assessment state is entered following 24 weeks of treatment with either fedratinib or BAT. Patients who discontinue treatment or die before reaching this state are labelled as an ‘early discontinuation’ or an ‘early death’. In the assessment state, patients undergo an instantaneous response assessment. The potential definitions of response used in the model are:

- Spleen response: ≥ 35% spleen volume reduction from baseline at 24 weeks
- Symptom response: ≥ 50% TSS reduction from baseline at 24 weeks
- Spleen or symptom response: ≥ 35% spleen volume or ≥ 50% TSS reduction from baseline at 24 weeks

In the base case, ‘spleen or symptom’ response is used to define response. This definition of response is based on the International Working Group for Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) guidelines,[51, 73] as defined in terms of either a spleen response or a symptom response. This approach has also been substantiated by the advisors attending the fedratinib Advisory Board. This is consistent with TA386.[7]

For fedratinib, time-to-discontinuation beyond 24 weeks is estimated separately for responders and non-responders. This is because non-responders discontinue treatment sooner (as observed in JAKARTA and JAKARTA-2). For BAT, an explicit time-to-discontinuation is not estimated; it is assumed the patient remains on BAT until another event due to the lack of alternative treatment options. There are other potential transitions (such as to progressed or end-of-life health states) that can occur in lieu of this explicit discontinuation.

The progressed state reflects transformation to acute myeloid leukaemia (AML). It was considered important to include secondary AML as a health state, to reflect its association with reduced life expectancy and health-related quality of life (HRQoL) in myelofibrosis.[74] Patients with AML do not return to the treatment states, but accrue any direct medical costs associated with AML (see Health-state unit costs and resource use).

End-of-life states in the model are palliative care and death. The palliative care health state reflects inpatient care in the final 8 weeks of life.[75, 76] Death is an absorbing health state.

B.3.2.2.3 Model implementation

To implement the DES approach, Visual Basic for Applications (VBA) in Microsoft Excel is used. VBA is the programming language of Microsoft Excel. The VBA code in the model adapts the example best-practice approach provided by the NICE Decision Support Unit (DSU).[77]

In the model, transitions occur according to time-to-discontinuation (TTD) curves, overall survival (OS) curves, time-to-AML curves, at response assessments, and

based on the proportion of patients expected to receive palliative care. The implementation of potential transitions within the model is summarised in Table 29.

Table 29: Implementation of events in the model

Event	Description	Assignment	Supporting data
Death (or early death)	Patient dies. 'Early death' is death before 24 weeks of treatment.	Time-to-death is sampled from parametric curves at the start of the simulation. No events can occur after death.	OS curves based on initial treatment
Early discontinuation	Patient discontinues treatment before 24 weeks.	When starting fedratinib, the proportion of early discontinuations input is used to determine who has this event. The time of discontinuation is assigned between 0 and 24 weeks of treatment, using a uniform distribution.	Trial data on the proportions of patients who discontinue before 24 weeks of fedratinib.
Response assessment	Patient undergoes response assessment at 24 weeks of treatment.	If a patient is receiving treatment at 24 weeks, they undergo a response assessment.	Structural model assumption informed by clinical trials.
Response	Patient classified as a responder to treatment.	The proportion of responders is used to determine who responds at 24 weeks. The denominator for response is corrected to consider only patients who reach the assessment (excludes early deaths and discontinuations). This ensures alignment between the % response input and the % response output.	ITCs were performed to adjust for imbalances between ITT response data for fedratinib in JAKARTA-2 and BAT in other trials (SIMPLIFY-2 and PERSIST-2).
Non-response	Patient classified as a non-responder to treatment. In scenario analysis, a stopping rule dictates that non-responders discontinue fedratinib treatment.	Patients who do not meet the criteria for response, as calculated above, are considered non-responders. If the stopping rule is used, the transition to BAT occurs immediately for these patients.	As above

Event	Description	Assignment	Supporting data
Loss of response	Patient originally classified as a responder loses response and its associated benefit. Patient remains on current treatment until discontinuation or another event.	A duration of response is sampled from a parametric curve. A patient loses response at that time or at discontinuation, whichever is earliest. In utility calculations, the patient only receives the responder utility increment while they respond.	Duration of response curves beyond 24 weeks (JAKARTA-2).
Discontinuation (fedratinib)	Patient stops receiving fedratinib.	At 24 weeks, a TTD is sampled from separate curves for responders and non-responders. If the patient reaches the TTD and is still on treatment, the patient will transition to BAT or palliative care, depending on whether they meet criteria explained below.	In the base case, TTD curves for responders and non-responders beyond 24 weeks are used. In scenario analysis, a stopping rule is enabled, and non-responders experience immediate discontinuation.
Worsening quality of life (BAT, *not used in base case*)	Patient receiving BAT experiences a worsening quality of life over time, independent of age-related utility decline.	The health state utility value applied to the patient in the BAT state is reduced every 24 weeks by a utility decrement. Only the utility changes, and the patient remains in the BAT health state.	The ruxolitinib appraisal (TA386) in a JAKi-naïve setting assumed that utility for patients on 'supportive care' (the last 30% of time on BAT) would fall every 24 weeks.
Progression to AML	Patient stops current treatment and enters AML health state.	A time-to-AML is assigned when the patient starts treatment, by sampling from a parametric curve. If a patient is alive at this time, they will progress to the AML health state when the time is reached. A new time-to-death (OS) is estimated upon progression, to reflect the reduced life expectancy associated with AML.	It is not clear whether treatment influences the rate of progression to AML. Therefore, the rate of progression to AML is set constant across treatments in the base case This can be based on the treatment and informed by respective trials.

Event	Description	Assignment	Supporting data
Transition to palliative care	Patient stops current treatment and enters palliative care health state.	For a patient on fedratinib, remaining life expectancy is assessed at the TTD. If there are ≤ 8 weeks of remaining life expectancy, they will move to palliative care. For a patient on BAT or in the AML state, no TTD is assigned. Therefore, a different rule is applied. In the final 8 weeks of life from the BAT or AML states, a proportion of patients are moved to palliative care.	Clinical assumptions based on premise that not all patients will receive palliative care, given that death is not always predictable.

Key: AML, acute myeloid leukaemia; BAT, best available therapy; ITT, intention-to-treat; JAK, Janus kinase; JAKi, JAK inhibitor; OS, overall survival; TA, technology appraisal; TTD, time-to-discontinuation.

B.3.2.2.4 Model features

Health effects in the model are calculated in life years (LYs) and QALYs.

Health effects and costs accrue for each patient based on their pathway over a lifetime time horizon. A 30-year time horizon was assumed to capture relevant outcomes over the lifetime of the patient, which was considered important as the model outcomes focus on survival. The average starting age in the model is approximately 66 years and 100% of patients are expected to die during the simulation. Shorter time horizons are explored in scenario analysis.

In the base case, the model considers a 3.5% annual discount rate for costs and QALYs in line with the NICE reference case. A 0% annual discount rate for LYs is used to align with standard practice.

Time cycles are not used in DES modelling. Therefore, considerations related to cycle length and half-cycle correction are not applicable.

Table 30: Features of the economic analysis

	Previous appraisals	Current appraisal	
Factor	**TA386**	**Chosen values**	**Justification**
Time horizon	Lifetime time horizon (30 years).	Lifetime time horizon (30 years).	The reference case stipulates that the time horizon should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. A lifetime time horizon is therefore considered sufficient to capture all meaningful differences.
Treatment waning effect?	None.	The implementation of *duration of response* within the model acts as a waning treatment effect, in that response is not artificially maintained for the entire treatment duration.	This reflects the clinical data to represent a more accurate portrayal of the disease. In a DES model, which uses a time-to-event framework, traditional hazards adjustment for treatment effect waning cannot be performed given there are no time cycles over which to do so. The same duration of response is used for both arms, which may be a conservative assumption given that a greater proportion of patients respond to fedratinib which may indicate deeper/longer response.
Source of utilities	A condition-specific preference-based measure for MF (MF-8D) was developed and applied to COMFORT-I data.	Treatment health state utilities were estimated using the MF-8D in JAKARTA-2. Other health state utilities (AML and palliative care) were externally sourced, and both estimated using the EQ-5D. A worsening utility decrement for BAT, applied in scenario	The NICE reference case stipulates that the EQ-5D is the preferred measure of health-related quality of life in adults. Some evidence suggests that the EQ-5D does not sufficiently capture HRQoL in myelofibrosis.[78] Therefore, the MF-8D, a condition-specific measure, was used where possible. Externally sourced utilities were used to appropriately estimate utilities that required longer-term data or greater sample size.

	Previous appraisals	Current appraisal	
		analysis, was taken from TA386.	The worsening utility decrement could be applied to BAT to reflect the worsening HRQoL assumed in TA386 for 'supportive care'. Clinical feedback indicated that BAT and 'supportive care' were equivalent.
Source of costs	Resource use unit costs were sourced from NHS Reference Costs and PSSRU Unit Costs. The main source for adverse event costs was a previous appraisal of Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel-containing regimen (TA316), which primarily used NHS Reference Costs. Administration costs were not included. Drug acquisition costs were taken from the BNF.	Resource use and adverse event cost sources were consistent with those used in TA386, using updated values or inflating values to a 2019 cost year. Administration costs were taken primarily from NHS Reference Costs. Drug acquisition costs were taken primarily from MIMS. eMIT was used for drugs available in generic form.	NHS Reference Costs, PSSRU, MIMS and eMIT are standard sources of UK-relevant costs and were used where possible. Where costs were not reported in these sources, cost inputs were sourced from appropriate literature.
Key: BNF, British National Formulary; eMIT, electronic medicines information tool; EQ-5D, EuroQol 5-dimensions questionnaire; NHS, National Health Service; MF-8D, myelofibrosis 8-dimensions; MIMS, Monthly Index of Medical Specialities; TA, technology appraisal; PSSRU, Personal and Social Services Research Unit.			

B.3.2.3 Intervention technology and comparators

Fedratinib was implemented in the model for the subset of patients in its marketing authorisation who have been treated with ruxolitinib. This population was chosen as it represents patients with significant unmet clinical need.

Fedratinib was modelled at a daily dose of 400 mg, administered orally. This is in line with the dose administered in the JAKARTA-2 trial and the expected fedratinib marketing authorisation.

The comparator in the model is best available therapy (BAT). BAT represents treatments received by a cohort of patients with MF, and is informed by the literature.[12, 15, 42] The use of BAT as a comparator aligns with the design of comparative clinical trials in MF and previous ruxolitinib economic modelling.

A 24-week treatment continuation rule for fedratinib can be implemented in the model. This reflects the stopping rule for ruxolitinib in TA386, by which patients whose disease did not respond to treatment were assumed to discontinue and receive BAT. The 24-week stopping rule and decision were based on the British Committee for Standards in Haematology guideline for the diagnosis and management of myelofibrosis (2012), which states that treatment should be stopped after 6 months if no reduction in splenomegaly has occurred or if symptoms have not improved since starting therapy.[79]

This submission assesses patients who have been treated with ruxolitinib and therefore require an alternative effective therapy. However, patients who stop ruxolitinib can rebound and lose symptom control.[40, 80, 81] Therefore, patients are often continued on ruxolitinib while achieving a suboptimal response, as no other targeted therapeutic options are available. It is for this reason that the BAT arms of Phase III trials in ruxolitinib-exposed populations have included substantial ruxolitinib use: from 44% in PERSIST-2 to 89% in SIMPLIFY-2[41, 42]. Clinical experts at a UK advisory board for fedratinib suggested that they would expect ruxolitinib use in the UK in this population to be similar to that used in SIMPLIFY-2.[18] Therefore, ruxolitinib alone is not considered a relevant comparator but is instead included as part of the

basket of treatments in BAT. This attempts to ensure alignment between costs and efficacy inputs (See Section B.3.6.1).

B.3.3. Clinical parameters and variables

B.3.3.1 Time-to-event data

In the model, patients are initialised with a scheduled set of event times. Example events could be treatment discontinuation or death. The event which occurs next in the model is the event with the lowest time-to-event. The occurrence of certain events will restrict other events from happening, e.g. death will end the simulation for a patient. For each patient, random numbers were used to sample times from time-to-event distributions (parametric curves).

Parametric curve fitting was conducted in line with NICE DSU guidance.[82] Six conventional parametric models were considered and compared: exponential, Gompertz, generalised gamma, log-logistic, log-normal, and Weibull. All curves were estimated using weeks as the unit of time and the 'flexsurvreg' R package. All curves were fit separately, meaning only data for the treatment of interest were considered in estimations. This was appropriate due to the non-randomised single-arm nature of the JAKARTA-2 trial.

Due to data immaturity, it was considered important to select the appropriate distributions informed primarily by the clinical plausibility of long-term predictions, remaining cognisant of statistical fit over the observed period.

B.3.3.2 Response

B.3.3.2.1 Types of response

Patient burden in myelofibrosis is linked to spleen size and symptom control.

Spleen response can be measured in terms of spleen volume or length. Spleen volume and length have considerable overlap. Evidence suggests that these measures produce equivalent results when assessed in a binary manner at established response thresholds (35% reduction in spleen volume, 50% reduction in spleen length).[83-85] This relationship was also observed in JAKARTA-2. A post-hoc

analysis of the JAKARTA-2 ITT population showed that spleen volume responders (n= 30, N = 97) were the same population as spleen length responders (n = 30, N = 97) by the above definitions. Only spleen volume response is carried forward into modelling, on the basis that spleen volume informed the primary efficacy endpoint of the JAKARTA and JAKARTA-2 trials.

In the JAKARTA studies, patients with spleen response underwent imaging 4 weeks after the 24-week assessment to confirm their response. This confirmatory step was not conducted in any other MF trials, and therefore response assessment at 24 weeks regardless of confirmation is used to enable like-for-like comparison between fedratinib and BAT.

Symptom response is usually patient-reported and summarised by the TSS from the Myelofibrosis Symptom Assessment Form (MF-SAF). An established threshold for symptom response is a 50% reduction in TSS.

International Working Group Myeloproliferative Neoplasms Research and Treatment guidelines suggest both types of response should be considered[51], this consideration was further substantiated at a clinical advisory board.[18] Acknowledging that both types of response are important, 'spleen or symptom' response classifies a patient as a responder for meeting either criterion.

B.3.3.2.2 Response assessment

In treatment health states, patients undergo a response assessment after 24 weeks of treatment (see Section B.3.2). Unlike TA386, the model in this submission allows patients on BAT to respond.

To support the comparison to BAT, given that JAKARTA-2 was a single-arm trial, indirect treatment comparisons (ITCs) were performed using response data for fedratinib in JAKARTA-2 and BAT in PERSIST-2 and SIMPLIFY-2 clinical trials.

In the economic model, the percentages of responders to fedratinib are aligned with the selection population, i.e. response percentages for the intermediate-2 and high-risk patients are used in the base case. However, the ITCs used the full ITT population data (or the Sensitivity cohort data in scenario analysis) to derive the

treatment effects. This is because PERSIST-2 and SIMPLIFY-2 clinical trials included significant proportions of intermediate-1 patients and it was not possible to separate this subgroup using the available data (Table 31).

The primary method of ITC was matching-adjusted indirect comparison (MAIC) which aims to address potential imbalances between trials on key prognostic variables with meaningful differences. The premise of MAIC methods is to adjust for between trial differences in baseline characteristics. When a common treatment comparator or 'linked network' is unavailable, a MAIC assumes that differences between absolute outcomes that would be observed in each trial are entirely explained by imbalances in prognostic variables and treatment effect modifiers. Under this assumption, every prognostic variable and every treatment effect modifier that is imbalanced between the two studies must be available.

Simulated treatment comparison (STC) has recently been discussed by NICE in their technical support document.[57] This approach was explored alongside the MAIC (and yielded highly similar results). For this method, a regression model is fitted to the JAKARTA-2 data for the outcome of interest, and should include all covariates that are prognostic or effect modifiers.[57] The model is then used to predict the percentage of fedratinib-treated patients who experience SVR or TSS reduction using the covariate values observed in the comparator evidence.[57] In addition, naïve ITCs between JAKARTA-2 and appropriate data sources were also performed to compare outcomes.

The ITCs were performed for spleen response, symptom response, and 'spleen or symptom' response. No appropriate comparator data were identified for the 'spleen or symptom' response definition, therefore a post-hoc analysis was performed by applying the minimum and maximum possible number of BAT 'spleen or symptom' responders from the available data to the ITC. The outcomes were used to produce an average result for this endpoint. Because of the importance of considering both spleen and symptom response as an endpoint, this analysis was the base case option used within the economic analysis. The ITC results for the separate spleen response and symptom response are presented in Appendix M.

An SLR was completed to identify studies from the literature that were suitable for performing an ITC. Two studies, PERSIST-2[41] and SIMPLIFY-2[42] were identified and are described alongside JAKARTA-2 in Table 31. The proportion of therapies received within BAT of the trial study arms is reported in Table 32. In both PERSIST-2 and SIMPLIFY-2, BAT mainly consisted of ruxolitinib. However, more patients in the SIMPLIFY-2 BAT (88.5%) arm received ruxolitinib compared with the PERSIST-2 (44.9%) BAT arm.

Table 31: Summary of study design and population in JAKARTA-2, PERSIST-2, and SIMPLIFY-2

	JAKARTA-2	**PERSIST-2**	**SIMPLIFY-2**
Phase	II	III	III
Design	Single-arm	RCT	RCT
Method of blinding	Open-label	Open-label	Open-label
Intervention (N)	Fedratinib 400 mg, once daily (starting dose) (97 [ITT])	Pacritinib 400 mg, once daily (75 [ITT efficacy population]) Pacritinib 200 mg, twice daily (74 [ITT efficacy population])	Momelotinib 200 mg once daily (104 [ITT])
Comparator (N)	NA	BAT (72 [ITT efficacy population])	BAT (52 [ITT])
Location	Multicentre	Multicentre	Multicentre
Method of randomisation	NA	1:1:1 ratio stratified by geographic region, risk category, and rebound platelet count	2:1 stratified by transfusion dependence and by baseline TSS
Crossover	NA	After Week 24 or progression of splenomegaly before Week 24	After completion of the randomised phase (24 weeks), all subjects were eligible to receive momelotinib in an extended treatment phase
Platelet count	≥ 50 x 10^9/L	≤ 100 x 10^9/L	There was no inclusion/exclusion criteria for platelet count at baseline

	JAKARTA-2	PERSIST-2	SIMPLIFY-2
DIPSS risk status:	N = 97	N = 72	N = 52
n (%) Intermediate-1	16 (16.5)[a]	13 (18.1)	16 (30.8)[b]
n (%) Intermediate-2	47 (48.5)	37 (51.4)	28 (53.8)
n (%) High-risk	34 (35.1)	22 (30.6)	8 (15.4)

Key: BAT, best available therapy; DIPSS, Dynamic International Prognostic Scoring System; ITT, intention-to-treat; N, number of subjects; NA, not applicable; RCT, randomised controlled trial; TSS, Total Symptom Score.
Note: [a], Intermediate-1 with symptoms; [b], Intermediate-1 with symptomatic splenomegaly or hepatomegaly

Table 32: BAT received in PERSIST-2 and SIMPLIFY-2

Study	PERSIST-2 (N=72)	SIMPLIFY-2* (N=52)
BAT received	Ruxolitinib (44.9%) Watch and wait (19.4%) Hydroxycarbamide (hydroxyurea) (19.4%) Prednisone (13.2%) Danazol (5.1%) Thalidomide (3.1%) Decitabine (2.0%) Interferon-alpha (2.0%)	Ruxolitinib (88.5%) Hydroxycarbamide (hydroxyurea) (23.1%) Corticosteroids (12.6%)

Key: BAT, best available therapy.
Notes: *, Only the most frequent treatments received were reported; the percentages in this table to do not sum to 100% as subjects could have received more than one therapy.

Table 33 presents spleen volume response from baseline to Week 24 for the ITT populations of the JAKARTA-2, PERSIST-2, and SIMPLIFY-2. The result for the subgroup of JAKARTA-2 patients with platelet count < 100 x 10^9/L has also been included for comparison to PERSIST-2 ITT results which only include patients with platelet count ≤ 100 x 10^9/L.

Table 33: Available data for subjects who have received prior ruxolitinib and achieved ≥ 35% spleen volume reduction

Outcome	JAKARTA-2[a]	PERSIST-2	SIMPLIFY-2
	FEDR 400 mg (N=97 [ITT][b])	**BAT (N=72 [ITT])**	**BAT (N=52 [ITT])**
≥ 35% SVR from baseline to Week 24/EOC 6 for the ITT population	30.9% (n=30, N=97)	NA[c]	6% (n=3, N=52)
≥ 35% SVR from baseline to Week 24 for subjects with platelet count < 100 x 10^9/L	[redacted]% (n=[redacted], N=33)	3%[d] (n=1, N=33)	NR

Key: BAT, best available therapy; EOC, End of Cycle; FEDR, fedratinib; ITT, intention-to-treat; n, number of responders; N, total patients; NA, not applicable; NR, not reported; SVR, spleen volume reduction.
Notes: [a], JAKARTA-2 includes patients with platelet count ≥ 50 x 10^9/L; the PERSIST-2 inclusion criteria included patients with a platelet count < 50 x 10^9/L; [b], ITT population includes all patients who received therapy, this is different to the MF-SAF population reported in Harrison et al[46]: [c], ITT results for PERSIST-2 include 53% of subjects who are JAK-inhibitor naïve; [d], results for the subgroup of subjects who had received prior ruxolitinib treatment.

Table 34 presents symptom response from baseline to Week 24 for the ITT populations of JAKARTA-2, PERSIST-2, and SIMPLIFY-2. As with spleen volume response, the result for the subgroup of JAKARTA-2 patients with platelet count < 100 x 10^9/L has also been included to facilitate comparison with the PERSIST-2 results which only included these patients.

Table 34: Available data for subjects who have received prior ruxolitinib and achieved ≥ 50% total symptom score reduction

Outcome	JAKARTA-2	PERSIST-2	SIMPLIFY-2
	FEDR 400 mg (N=97 [ITT] [a])	**BAT (N=72 [ITT])**	**BAT (N=52 [ITT])**
≥ 50% reduction in TSS from baseline to 24 weeks for the ITT population	[redacted]% (n=[redacted]; N=97)	NA[b]	6% (n=3; N=51)
≥ 50% reduction in TSS from baseline to 24 weeks for the subjects with platelet count < 100 x 10^9/L	[redacted]% (n=[redacted]; N=33)	15%[c] (n=5; N=33)	NR

Key: BAT, best available therapy; FEDR, fedratinib; ITT, intention-to-treat; n, number of responders; N, total patients; NR, not reported; TSS, total symptom score.
Notes: [a], ITT population includes all patients who received therapy, this is different to the MF-SAF population reported in Harrison et al[46]; [b], ITT results for PERSIST-2 include 53% of subjects who are JAK-inhibitor naïve; [c], results for the subgroup of subjects who had received prior ruxolitinib treatment.

B.3.3.2.3 Indirect treatment comparison (ITC) methods

MAICs and STCs were performed using the BAT-arm data from SIMPLIFY-2,[42] whereas only a naïve ITC could be performed with the BAT-arm data from PERSIST-2[41] as baseline characteristics were not reported for the ruxolitinib-treated subgroup. Identification of prognostic factors was therefore only limited to the JAKARTA-2 and SIMPLIFY-2 studies. In published literature, treatment effect modifiers could not be identified for JAK-inhibitor exposed patients. Furthermore, exploratory analyses to identify treatment effect modifiers using the JAKARTA-2 patient level data was not possible given that JAKARTA-2 is a single-arm trial. Instead, statistical analysis supported by further input from an external haematologist was performed and is described below.

Univariate models for spleen volume reduction (SVR) and TSS reduction were fitted for each of the available patient characteristics. The p-values for both SVR and TSS analyses are reported in Appendix L. For each variable, patients with missing information were removed, and a likelihood-ratio test was performed to understand the significance of the variable on SVR and TSS reduction. Due to multiple testing, interpretation of p-values was made with caution. In the JAKARTA-2 patient-level

data, transfusion dependence was a ‘perfect predictor’ of SVR. In logistic regression, when a variable is a perfect predictor of the outcome, the model struggles to converge. This is a problem that is referred to as complete separation.[58] Therefore, for SVR, the p-value from the univariate analysis of transfusion dependence was interpreted with further caution.

In addition to the univariate analysis, variables were compared to determine whether there was significantly high standardised difference between the JAKARTA-2 and SIMPLIFY-2 trial data. All variables, with the exception of MF subtype and baseline TSS, had a standardised difference greater than 10%, indicating potential imbalance. A multivariate analysis was also completed with all variables to determine which were significant in forward selection by Akaike Information Criterion (AIC).

An external haematologist involved and experienced in developing MF prognostic risk scores and clinical trials of novel MF treatments was consulted to acquire further input on which imbalances between baseline characteristics in JAKARTA-2 and SIMPLIFY-2 could be considered clinically meaningful.[86] Specifically, the external haematologist was presented with the reported baseline characteristics of the BAT-treated patients in SIMPLIFY-2 and corresponding baseline characteristics of the fedratinib-treated patients in JAKARTA-2 in a table. The haematologist was asked to indicate whether the magnitude of the differences observed for each baseline characteristic would be a driver of differences in achieving ≥ 35% SVR beyond that of the treatment itself. Subsequently, three characteristics were identified as having differences that are clinically meaningful: ECOG PS, DIPSS, and transfusion dependence status.

Prognostic factors that were used to adjust the indirect treatment comparisons were included if they satisfied both the following criteria:

- The variable was identified as having clinically meaningful imbalance by an external haematologist
- The variable was also identified as being an important prognostic factor in the JAKARTA-2 study (from either the univariate or multivariable analyses)

Variables fulfilling both criteria for SVR were ECOG PS and transfusion dependence and variables fulfilling both criteria for TSS reduction were ECOG PS and DIPSS. These variables were therefore explored in the respective adjusted analyses as presented in Appendix L. These prognostic factors were then combined to use within the 'spleen or symptom' analysis to produce the base case response estimates.

B.3.3.2.4 Spleen or symptom response results

The 'spleen or symptom' ITC was performed between the available BAT and fedratinib trial data with the endpoint of patients having experienced either SVR or TSS. Given the limitation that the available SIMPLIFY-2 and PERSIST-2 data for the BAT arm does not report this endpoint, the analyses are performed with one of two assumptions:

- The number of BAT patients who reach the endpoint is equal to the maximum number of patients experiencing either SVR or TSS response separately – Referred to henceforth as the Minimum BAT response scenario.
- The number of BAT patients who reach the endpoint is equal to the sum of patients experiencing either SVR or TSS response separately – Referred to henceforth as the Maximum BAT response scenario

Using these assumptions, Table 35 and Table 36 summarise the comparison scenarios that were made in this ITC analysis.

Table 35: Summary of comparisons to the PERSIST-2 evidence

Comparison made	Data used to make the comparison	
	JAKARTA-2: FEDR 400 mg (N=97 [ITT][a])	**PERSIST-2: BAT (N=72 [ITT])**
Minimum BAT response scenario – Unadjusted ITC of the proportion of ITT subjects achieving either ≥ 35% SVR from baseline to Week 24/EOC 6 OR ≥ 50% TSS reduction from baseline to Week 24/EOC 6 in ITT subjects with platelet count < 100 x 10^9/L	[redacted]% (n=[redacted]; N=33)	15.2% (n=5; N=33)
Maximum BAT response scenario - Unadjusted ITC of the proportion of ITT subjects	[redacted]% (n=[redacted]; N=33)	18.2% (n=6; N=33)

Company evidence submission template for fedratinib for splenomegaly and symptoms in myelofibrosis ID1501

Comparison made	Data used to make the comparison	
	JAKARTA-2: FEDR 400 mg (N=97 [ITT][a])	**PERSIST-2: BAT (N=72 [ITT])**
achieving either ≥ 35% SVR from baseline to Week 24/EOC 6 OR ≥ 50% TSS reduction from baseline to Week 24/EOC 6 in ITT subjects with platelet count < 100 x 10^9/L		
Key: BAT, best available therapy; EOC, End of Cycle; FEDR, fedratinib; ITC, indirect treatment comparison; ITT, intention-to-treat; n, number of responders; N, total number of subjects; SVR, spleen volume reduction; TSS, Total Symptom Score; wo, without. **Notes:** [a], ITT population includes all patients who received therapy, this is different to the MF-SAF population reported in Harrison et al[46]:		

Table 36: Summary of comparisons to the SIMPLIFY-2 evidence

Comparisons made	Data used to make the comparison	
	JAKARTA-2: FEDR 400 mg (N=97 [ITT][a])	**SIMPLIFY-2: BAT (N=52 [ITT])**
Minimum BAT response scenario - ITCs (unadjusted, MAICs, and STCs) of the proportion of ITT subjects achieving either ≥ 35% SVR from baseline to Week 24/EOC 6 OR ≥ 50% TSS reduction from baseline to Week 24/EOC 6	[redacted]% (n=[redacted]; N=97)	5.8% (n=3; N=52)
Maximum BAT response scenario - ITCs (unadjusted, MAICs, and STCs) of the proportion of ITT subjects achieving either ≥ 35% SVR from baseline to Week 24/EOC 6 OR ≥ 50% TSS reduction from baseline to Week 24/EOC 6	[redacted]% (n=[redacted]; N=97)	11.5% (n=6; N=52)
Key: BAT, best available therapy; EOC, End of Cycle; FEDR, fedratinib; ITC, indirect treatment comparison; ITT, intention-to-treat; MAIC, matching-adjusted indirect treatment comparison; n, number of responders; N, total number of subjects; STCs, simulated treatment comparisons; SVR, spleen volume reduction; TSS, Total Symptom Score; wo, without. **Notes:** [a], ITT population includes all patients who received therapy, this is different to the MF-SAF population reported in Harrison et al[46]:		

There were three JAKARTA-2 patients with missing ECOG PS information at baseline. It was therefore assumed that these patients had an ECOG PS of either 0

or 1 (rather than ECOG PS 2) to calculate the matching weights as this was the most prevalent group. Furthermore, three patients in SIMPLIFY-2 had missing information on whether they were transfusion dependent or independent at baseline. It was assumed that these patients were distributed equally across the two categories; the reported percentages were used to represent the entire SIMPLIFY-2 BAT population.

Following the matching procedure (weighting on ECOG PS, DIPSS and transfusion dependence), the weighted baseline characteristics for JAKARTA-2 patients were compared with the comparator population (the SIMPLIFY-2 BAT arm). Table 37 indicates that the MAIC method led to reweighted JAKARTA-2 covariates that are the same as the SIMPLIFY-2 population. However, the matching procedure led to a relatively small effective sample size (ESS) for the JAKARTA-2 population (ESS was 34.4 compared to the original sample size of 97). It was noted that there was a large imbalance in the proportion of patients who were transfusion dependent between the fedratinib-treated patients and BAT-treated patients and removal of transfusion dependence resulted in an ESS value of 81.6. Because of this, an additional analysis using only adjustment for ECOG PS and DIPSS was included in the MAIC.

Table 37: Sample size/effective sample size and baseline characteristics before and after matching

	N/ESS	**ECOG PS: % 0 or 1**	**DIPSS: % Intermediate-1 or 2**	**% Transfusion dependent**
SIMPLIFY-2 BAT arm	N=52	86.5	84.6	51.9
JAKARTA-2 population before matching	N=97	76.3	64.9	14.4
JAKARTA-2 population after matching on ECOG PS, DIPSS and transfusion dependence	ESS=18.3 (35.5% of original sample size [N=97])	86.5	84.6	51.9

	N/ESS	ECOG PS: % 0 or 1	DIPSS: % Intermediate-1 or 2	% Transfusion dependent
JAKARTA-2 population after matching on ECOG PS and DIPSS	ESS=81.6 (84.1% of original sample size [N=97])	86.5	84.6	NA
Key: BAT, best available therapy; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ESS, effective sample size; N, number of subjects; NA, not applicable; SVR, spleen volume reduction.				

Minimum BAT response scenario results

Table 38 presents the naïve ITC results for the proportion of patients achieving ≥ 35% SVR or ≥ 50% TSS reduction from baseline to Week 24 (spleen or symptom response) when the BAT evidence is informed by the PERSIST-2 data, using the minimum BAT response assumption.

Fedratinib 400 mg led to a greater proportion of patients achieving spleen or symptom response compared to BAT. Fedratinib 400 mg had a [redacted]% (95% confidence interval [CI] [redacted]) greater proportion of patients with spleen or symptom response compared with BAT.

Table 38: Fedratinib 400 mg versus BAT (from PERSIST-2) – naïve ITC results for the spleen volume reduction or total symptom score endpoint (minimum BAT response scenario)

	JAKARTA 2 (400 mg FEDR) N=33	PERSIST-2 (BAT) N=33
≥ 35% SVR or ≥ 50% TSS from baseline to Week 24 (subgroup of the JAKARTA-2 ITT population with platelet counts < 100 x 10^9/L and without LOCF)	[redacted]% (n=[redacted])	15.2% (n=5)
	Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [[redacted] [redacted]]	
Key: BAT, best available therapy; CI, confidence interval; FEDR, fedratinib; ITC, indirect treatment comparison; ITT, intention-to-treat; n, number of responders; N, total number of subjects; SVR, spleen volume reduction; TSS, total symptom score		

The ITC results when the BAT evidence is informed by the SIMPLIFY-2 data, using the minimum BAT response assumption, are presented in Table 39. All analysis (unadjusted, MAIC, STC) showed that fedratinib 400 mg consistently led to a greater proportion of patients achieving spleen or symptom response compared to BAT. When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with spleen or symptom response compared with BAT. Using MAIC methods, after adjustment for baseline ECOG PS and DIPSS the treatment effect decreased slightly to [redacted]% (95% CI: [redacted]). After adjustment for baseline ECOG PS, DIPSS and transfusion dependence, fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of responders than BAT. However, as discussed above, the results with adjustment for ECOG PS and transfusion dependence should be interpreted with caution given the relatively small effective sample size.

The STC analyses are only presented for the adjustment of ECOG PS and DIPSS as the adjustment for ECOG PS, DIPSS and transfusion dependence resulted in a model that had a very large standard error for the transfusion dependence coefficient (1,818.1 – likely due to complete separation). For the adjustment of ECOG PS and DIPSS, the MAIC and STC methods generated similar results; treatment effects were [redacted]% (95% CI: [redacted]) and [redacted]% (95% CI: [redacted]) using the MAIC and STC methods, respectively

Table 39: Fedratinib 400 mg versus BAT (from SIMPLIFY-2) – unadjusted and adjusted ITC results for the SVR or TSS endpoint (minimum BAT response scenario)

Method	Variables included in adjustment	JAKARTA 2 (400 mg FEDR)	SIMPLIFY-2 (BAT)
Unadjusted ITC	• NA	[redacted]% (n=[redacted]; N=97)	5.8% (n=3; N=52)
		Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [[redacted]]	

Method	Variables included in adjustment	JAKARTA 2 (400 mg FEDR)	SIMPLIFY-2 (BAT)
MAIC	• ECOG PS • DIPSS	[redacted]% (CI: [redacted])[a]	5.8% (n=3; N=52)
		Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [[redacted]][a]	
MAIC	• ECOG PS • DIPSS • Transfusion dependence	[redacted]% (CI: [redacted])[a]	5.8% (n=3; N=52)
		Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [[redacted]][a]	
STC	• ECOG PS • DIPSS	[redacted]% (CI: [redacted])	5.8% (n=3; N=52)
		Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [[redacted]]	

Key: BAT, best available therapy; CI, confidence interval, DIPSS, Dynamic International Prognostic Scoring System; ECOG PS, Eastern Cooperative Oncology Group Performance Status; FEDR, fedratinib; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; n, number of responders; N, total number of subjects; NA, not applicable; STC, simulated treatment comparison.
Note: [a]Bootstrap percentile CI (based on 10,000 samples).

Maximum BAT response scenario results

Table 40 presents the naïve ITC results when the BAT evidence is informed by the PERSIST-2 data and the maximum BAT response assumption is applied.

Fedratinib 400 mg led to a [redacted]% (95% CI: [redacted]) greater proportion of patients achieving spleen or symptom response compared to BAT.

Table 40: Fedratinib 400 mg versus BAT (from PERSIST-2) – naïve ITC results for the spleen volume reduction or total symptom score endpoint (maximum BAT response scenario)

	JAKARTA 2 (400 mg FEDR) N=33	**PERSIST-2 (BAT) N=33**
≥ 50% reduction in TSS from baseline to Week 24 (subgroup of the JAKARTA-2 ITT population with platelet counts < 100 x 10^9/L and without LOCF)	[redacted]% (n=20)	18.2% (n=6)
	Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [redacted]	
Key: BAT, best available therapy; CI, confidence interval; FEDR, fedratinib; ITC, indirect treatment comparison; ITT, intention-to-treat; LOCF, last observation carried forward; n, number of responders; N, total number of subjects; SVR, spleen volume reduction; TSS, total symptom score		

Under the maximum BAT response assumption, the ITC results when the BAT evidence is informed by the SIMPLIFY-2 data are presented in Table 41. Again, in all analyses fedratinib 400 mg consistently led to a greater proportion of patients achieving spleen or symptom response compared to BAT. When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of spleen or symptom responders compared with BAT. The MAIC, which adjusted for ECOG PS and DIPSS, indicated a treatment effect of [redacted]% (95% CI: [redacted]).

The MAIC and STC analyses produced similar results after adjustment for ECOG PS and DIPSS; with treatment effects of [redacted]% (95% CI: [redacted]) and [redacted]% (95% CI: [redacted]) using MAIC and STC methods, respectively.

Table 41: Fedratinib 400 mg versus BAT (from SIMPLIFY-2) – unadjusted and adjusted ITC results for the spleen volume reduction or total symptom score endpoint (maximum BAT response scenario)

Method	Variables included in adjustment	JAKARTA 2 (400 mg FEDR)	SIMPLIFY-2 (BAT)
Unadjusted ITC	• NA	[redacted]% (n=[redacted]; N=97)	11.5% (n=6; N=52)
		Δ 400 mg Fed–BAT [95% CI]: [redacted]% [[redacted]]	
MAIC	• ECOG PS • DIPSS	[redacted]% ([redacted])[a]	11.5% (n=6; N=52)
		Δ 400 mg Fed–BAT [95% CI]: [redacted]% [[redacted]]	
MAIC	• ECOG PS • DIPSS • Transfusion dependence	[redacted]% (CI: [redacted])[a]	11.5% (n=6; N=52)
		Δ 400 mg FEDR–BAT [95% CI]: [redacted]% [[redacted]][a]	
STC	• ECOG PS • DIPSS	[redacted]% (CI: [redacted])	11.5% (n=6; N=52)
		Δ 400 mg Fed–BAT [95% CI]: [redacted]% [[redacted]]	

Key: BAT, best available therapy; CI, confidence interval, DIPSS, Dynamic International Prognostic Scoring System; ECOG PS, Eastern Cooperative Oncology Group Performance Status; FEDR, fedratinib; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; n, number of responders; N, total number of subjects; NA, not applicable; STC, simulated treatment comparison.
Note: [a]Bootstrap percentile CI (based on 10,000 samples).

Summary

The MAIC, STC and naïve analyses within the maximum and minimum BAT response scenarios produced similar results. The base case analysis used within the model was the MAIC analysis using the ECOG PS and DIPSS scores for adjustment, given the relatively high effective sample size and given that no issues were experienced with convergence (which occurred in some STC analyses). The treatment effects from the minimum and maximum BAT response analyses provided

a range within which the expected treatment effect would lie. Therefore, these analyses were used to produce an average treatment effect (Table 42). The active fedratinib response percentage for the SVR or TSS endpoint was taken from the JAKARTA-2 Int-2/High-risk base case population (Table 43) and the response adjustment percentages were applied to produce the active response proportions for each therapy arm within the model (Table 44).

Table 42: Application of response adjustments in base case

Treatment	% difference to fedratinib	% difference to fedratinib (upper bound)	% difference to fedratinib (lower bound)
BAT (minimum response)	[redacted]		
BAT (maximum response)	[redacted]		
BAT (used)	[redacted]	[redacted]	[redacted]

Table 43: JAKARTA-2 Int-2/High-risk spleen or symptom response

Treatment	n	N	%	Source
Fedratinib - JAKARTA-2 ITT - intermediate-2+ risk	[redacted]	81	[redacted]	JAKARTA-2 post-hoc analysis, intermediate-2 or high risk

Table 44: Base case response probabilities at 24 weeks

Treatment	Active Probability
Fedratinib	[redacted]
Best available therapy	[redacted]

When the active endpoint is switched to either of the separate SVR or TSS response analyses results (Appendix L) then the percentage adjustment is applied to the active fedratinib percentage in the same way, without the requirement to take an average from maximum and minimum BAT scenario analyses.

The Sensitivity cohort was also analysed for 'spleen or symptom' response in the same way at the ITT population. These results are presented in Appendix L.

B.3.3.2.5 Duration of response

Patients can 'lose' response on fedratinib but remain on treatment, because some clinical benefit may continue for the patient which does not meet the criteria for complete discontinuation.

The model accounts for this by estimating a duration of response (DoR). After response is lost, the utility increment associated with response is lost, and patients instead experience the utility increment associated with non-response. Therefore, response is not artificially maintained for the entire treatment duration, which aims to reflect clinical practice. This implementation of waning was not modelled for ruxolitinib in TA386.

Duration of response in the model was based on spleen response, as DoR data for other response definitions were not collected for fedratinib or available for other treatments.

In the JAKARTA studies, duration of spleen response was defined as the time from the date of the first Independent Review Committee (IRC)-assessed response (≥ 35% spleen volume reduction) to the date of subsequent IRC-assessed progressed disease or death, whichever was earlier. Parametric curves were fitted to the DoR data for fedratinib in JAKARTA-2 for the Int-2/High-risk population, with the ITT population fitted as an option for scenario analysis. The analysis was performed for spleen responders only, and the number of patients at risk (Int-2/High-risk n = [redacted], ITT n = 30) for this outcome fell over time partly due to censoring (Figure 16). For this reason, the generalised gamma curve failed to converge for either population. All curves gave a similar extrapolation, apart from the exponential curve which predicted that patients would respond for significantly longer, and so was excluded.

Of the remaining curves, the log-normal curve was chosen as it performed best in terms of statistical fit (AIC and Bayesian Information Criterion [BIC]). Table 45 presents AIC and BIC values for the base case analysis.

Figure 16: Parametric curves fit to duration of response data in JAKARTA-2 Int-2/High-risk population

Key: KM, Kaplan–Meier.

Table 45: Statistical fit of duration of response curves in JAKARTA-2 Int-2/High-risk population

Distribution	AIC	BIC
Exponential	28.50	29.79
Generalised gamma	-	-
Gompertz	26.78	29.37
Log-logistic	25.47	28.06
Log-normal †	**25.03**	**27.63**
Weibull	25.82	28.41

Key: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
Notes: †, Selected distribution.
Values in bold indicate best fitting parametric fit. The generalised gamma model did not converge so fit statistics could not be derived.

B.3.3.3 Overall survival

In the base case for the model, OS is estimated based on an OS curve for the first treatment received by the patient.

The JAKARTA-2 trial only assessed fedratinib. Therefore, external sources of OS data were considered for BAT. The searches performed for the clinical SLR (described in Appendix D) were first conducted in August 2018, and subsequently updated to inform this submission. In February 2019, the August 2018 iteration of the clinical SLR was updated systematically using Embase to identify overall survival evidence for patients after discontinuation of ruxolitinib. The review retrieved 4,011 publications, of which 11 reported survival for the population of interest.[7, 16, 41, 87-94] Following the review, two further relevant studies were published: Schain et al. (2019) and Palandri et al. (2019).[15, 17] The latter provided updated data and information from the Palandri et al. (2018) study from the original review.[93] Therefore, 13 studies are summarised in Table 46 and Table 47.

Table 46: Characteristics of studies reporting overall survival after ruxolitinib discontinuation

Study name	Study design	Publication format	Patient numbers	Treatment received after ruxolitinib discontinuation
Investigational agents				
Mascarenhas et al., 2018 (a) [87]	Randomised Phase II study	Abstract	48 (4.7 mg/kg) and 57 (9.4 mg/kg)	Imetelstat 4.7 mg/kg, and 9.4 mg/kg
Mascarenhas et al., 2018 (b) [41]	Randomised Phase III study	Manuscript	100	Pacritinib 200 mg, pacritinib 400 mg, and BAT
Best available therapy or approved treatment				
Gupta et al., 2016 [88]	Two-stage Simon	Abstract	21	Ruxolitinib + ASCT
Kadir et al., 2018 [89]	Retrospective observational	Manuscript	171	ASCT
Kuykendall et al., 2017 [16]	Retrospective observational	Manuscript	22	No treatment
			25	Salvage therapy (lenalidomide, thalidomide, hydroxycarbamide, interferon, danazol, hypomethylating agents, investigational agents)
			63	All patients (no treatment, salvage therapy, ASCT)
Mehra et al., 2016 [90]	Retrospective observational (claims database)	Abstract	63	Non-ruxolitinib treatment
			488	2L-ruxolitinib
Miller et al., 2018 [91]	Retrospective observational	Abstract	41	ASCT

Study name	Study design	Publication format	Patient numbers	Treatment received after ruxolitinib discontinuation
Newberry et al., 2017 [92]	Non-randomised study	Manuscript	56	Hydroxycarbamide, investigational agents, splenectomy, ASCT, hypomethylating agents, induction chemotherapy, anagrelide
NICE (COMFORT-II), 2016 [7]	Randomised Phase III study	HTA	39	NR
Palandri et al., 2018 [93]	Retrospective observational	Abstract	NR	Evaluable population
			NR	Conventional agents (including hydroxycarbamide, danazol, anagrelide, ESA)
Shanavas et al., 2016 [94]	Retrospective observational	Manuscript	100	ASCT
Sourced after review				
Palandri et al., 2019 [17]	Retrospective observational	Manuscript	218	Conventional agents; novel agents (JAK-inhibitors, imetelstat, PRM-151)
Schain et al., 2019 [15]	Retrospective observational	Manuscript	71	Conventional agents (including glucocorticoids, hydroxycarbamide)

Key: 2L, second line; ASCT, autologous stem cell transplantation; BAT, best active treatment; CI, confidence interval; ESA, erythropoiesis-stimulating agents; HR, hazard ratio; HTA, health technology assessment; JAK, Janus kinase; KM, Kaplan–Meier; kg, kilogram; mg, milligram; N, number of patients; NR, not reported; OS, overall survival.

Table 47: Survival outcomes in studies reporting overall survival after ruxolitinib discontinuation

Study name	Survival outcomes
Investigational agents	
Mascarenhas et al., 2018 (a) [87]	Imetelstat: Median OS: 19.9 months (4.7 mg/kg) and 29.9 months (9.4 mg/kg)
Mascarenhas et al., 2018 (b) [41]	HRs relative to BAT in a JAKi-exposed subgroup: Pacritinib 200 mg: 0.49 (95% CI: 0.12–1.96) Pacritinib 400 mg: 1.80 (95% CI: 0.62–5.23)
Best available therapy or approved treatment	
Gupta et al., 2016 [88]	In all 21 patients from time of registration with median follow-up of 5.8 months, 6-month OS was 75% (95% CI 44–90%) before transplant in prior JAKi exposed patients. For the 19 transplant recipients, 6-month OS was 83% (95% CI 55–94%) from date of transplant in prior JAKi exposed patients.
Kadir et al., 2018 [89]	OS rate: ruxolitinib + ASCT vs non-ruxolitinib + ASCT: 72.7% vs 69.9%; *P* = 0.4
Kuykendall et al., 2017 [16]	No treatment: median OS: 4.9 months
	All patients: median OS: 13.0 months
	Salvage therapy: median OS: 15.0 months
Mehra et al., 2016 [90]	Non-ruxolitinib treatment: median OS: 14 months
	2L-ruxolitinib: median OS reported as 30 months, although there were 0 patients at risk at 30 months. Suggest interpretation with caution.
Miller et al., 2018 [91]	ASCT with prior ruxolitinib vs without prior ruxolitinib: HR = 0.53 (95% CI: 0.26–1.07); *P* = 0.077
Newberry et al., 2017 [92]	Median OS: 14 months
NICE (COMFORT-II), 2016 [7]	Median OS: 16 months (read from Kaplan–Meier data) in ‘early discontinuers’ and ‘spleen responders’

Study name	Survival outcomes
Palandri et al., 2018 [93] *(see below for updated data)*	Evaluable population: median OS: 22.6 months
	Conventional agents: median OS: 30 months (patients who discontinued in chronic phase) Novel agents: median OS: not reached at 40 months *(patients who discontinued in chronic phase)*
Shanavas et al., 2016 [94]	The 2-year OS probability: 61% (95% CI: 49–71).
Sourced after review:	
Palandri et al., 2019 [17]	Overall: median OS: 13.2 months *(all patients)* Conventional therapies: median OS: 28.9 months *(patients who discontinued in chronic phase)* Novel agents: median OS: 40.5 months (patients who discontinued in chronic phase)
Schain et al., 2019 [15]	Overall: median OS: 16 months
Key: 2L, second line; ASCT, autologous stem cell transplantation; BAT, best active treatment; CI, confidence interval; ESA, erythropoiesis-stimulating agents; HR, hazard ratio; HTA, health technology assessment; JAKi, Janus kinase inhibitor; KM, Kaplan–Meier; kg, kilogram; mg, milligram; N, number of patients; NR, not reported; OS, overall survival.	

Of the 13 included studies, eight were retrospective observational studies,[15-17, 89-91, 93, 94] three were randomised controlled trials (RCTs),[7, 41, 87] one was a non-RCT,[92] and one was a two-stage Simon study.[88] Across all the included studies, eight reported median OS,[7, 15-17, 87, 90, 92, 93], with variation due to disease status and type of treatments received after ruxolitinib discontinuation.

Patients receiving observation or no treatment after ruxolitinib had a median OS of 4.9 months.[16] Median OS in patients who received treatment with salvage therapy or conventional agents (e.g. hydroxycarbamide, danazol, anagrelide) was typically around 14–15 months.[16, 90, 92] In general, estimates of median OS for whole study populations were typically between 13–16 months.[15-17, 92] This is in line with the median OS following ruxolitinib discontinuation in the COMFORT-II study, which was approximately 16 months (read from a KM plot in NICE TA386), but was only reported for early discontinuers and spleen responders.[7]

A retrospective analysis of European registry data estimated median OS for patients with MF receiving novel agents (such as fedratinib) was 40.5 months, while median OS for conventional agents after ruxolitinib was 28.9 months.[17] The comparatively higher survival observed in this study is likely due to the inclusion of intermediate-1 risk patients and reporting which excludes patients in the ‘blast phase’ of myelofibrosis.

The OS data from four external studies were included as options in the model, based on providing potentially representative and relevant estimates of survival for patients receiving BAT following ruxolitinib. Additionally, these studies reported KM plots for the population of interest, such that they could be digitised to create pseudo-patient level data.

The four included studies were:

1. Schain et al., 2019[15]
2. COMFORT-II (spleen responders or early discontinuations), as reported in TA386[7]
3. Kuykendall et al., 2017[16]
4. Palandri et al., 2019[17]

The KM plots are presented in Figure 17 compared against the JAKARTA-2 OS data. No study provided sufficient baseline characteristics specific to the digitised KM population to allow for an adjusted ITC for OS (Appendix Section L.5.1)

Figure 17: Available KM data for post-ruxolitinib survival

Key: KM, Kaplan–Meier

Prior to the UK advisory board held for fedratinib, clinician attendees (N=7) were asked to consider and provide their expectations of survival in the post ruxolitinib population for those treated with BAT and those treated with fedratinib. Attendees provided estimates at the following time points: 1 year, 2 years, 5 years, 10 years, 15 years, and 20 years. The averages of these estimates are shown in Table 48. At each timepoint, the experts suggested that a higher proportion of patients would be alive having received fedratinib instead of current BAT.

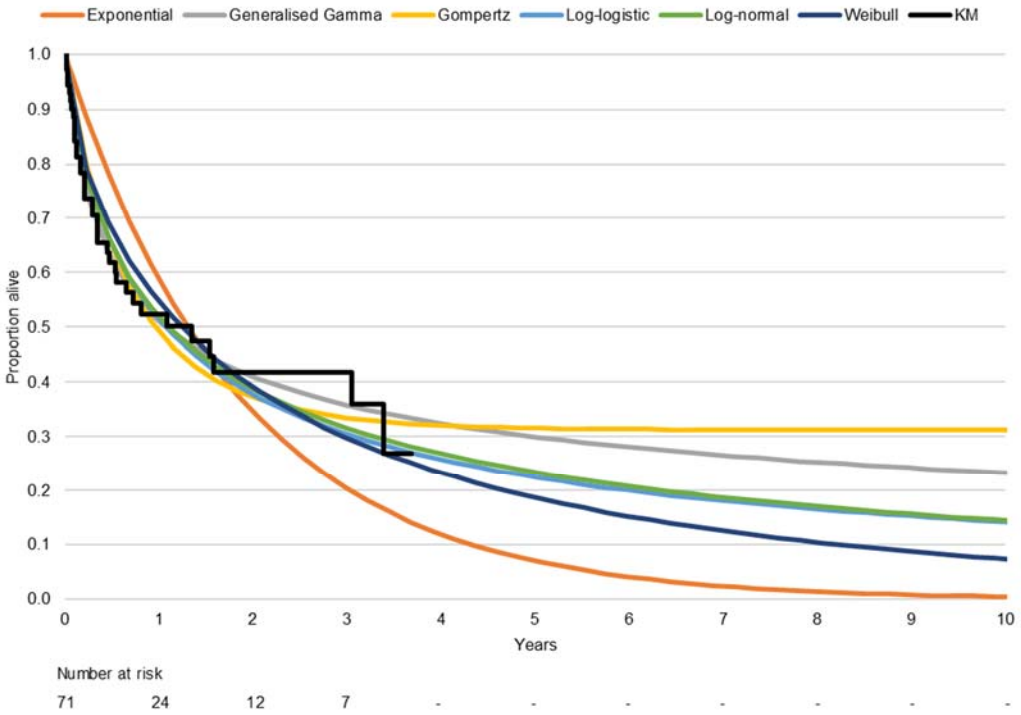
Table 48: Patients alive over time post ruxolitinib (average estimates taken pre-advisory board)

Treatment	1 year	2 years	5 years	10 years	15 years	20 years
BAT	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Fedratinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Key: BAT, best available therapy.						

At the advisory board, clinicians were shown summaries for each of the four potential OS sources for BAT. Summaries included information on the patient population, sample size, median survival, risk status of patients, study design, treatment composition, and time of last observation. For each source, KM plots with numbers at risk were presented. Parametric curves were then added alongside information on predicted survival (median, mean, and proportion alive at each time point listed above), and statistical fit (AIC and BIC).

The group indicated that the population most representative of those expected to receive fedratinib in UK practice was that of Schain et al (2019). The risk status of patients was not recorded within the Schain et al study, although it is assumed all patients would have a risk status of intermediate-2 or above given the approval of ruxolitinib use in Norway and Sweden.[15] Of the 6 parametric curves for this source (Figure 18), the group indicated that the exponential and Weibull were most relevant and representative of UK patients. The Weibull curve was selected in the base case as it provided a better statistical fit to the data.

Figure 18: Parametric curves fit to overall survival data in Schain et al., 2019



Key: KM, Kaplan–Meier.

Schain et al. (2019) assessed the survival of 190 patients in MF patients treated with ruxolitinib both at treatment initiation and post-discontinuation in Norway and Sweden.[15] Survival data for 71 patients who discontinued ruxolitinib were reported. Median survival in these patients was 16 months. The most common treatment received following ruxolitinib was glucocorticoids (65.9%) followed by hydroxycarbamide (32.4%).[15] However, a small proportion of medicines used in the Schain population may not reflect UK practice, e.g. thalidomide (5%). The statistical fit of the parametric curves is presented in Table 49.

Table 49: Statistical fit of overall survival curves in Schain et al., 2019

Distribution	AIC	BIC
Exponential	404.42	406.69
Generalised gamma	**387.36**	394.15
Gompertz	391.56	396.09
Log-logistic	389.88	394.40
Log-normal	387.51	**392.03**
Weibull†	393.07	397.59

Key: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
Notes: †, Selected distribution.
Values in bold indicate best fitting parametric fit.

For fedratinib, the clinicians were shown parametric curves fit to the survival data for the JAKARTA-2 ITT population (N=97). During the meeting it was advised that only the intermediate-2 and high-risk population would receive fedratinib in the UK; therefore, expectations of survival were provided with this in mind. Figure 19 shows the similarity between the KMs of the ITT and intermediate-2 and high-risk populations. Figure 20 shows the parametric curves as presented at the advisory board.

Figure 19: JAKARTA-2 overall survival data (comparison of populations)

Key: Int2/HR, intermediate-2 and high-risk subgroup; ITT, intention-to-treat; KM, Kaplan–Meier

Figure 20: Parametric curves fit to overall survival data in the JAKARTA-2 ITT population

Key: ITT, intention-to-treat; KM, Kaplan–Meier.

The clinicians indicated that both the exponential and Weibull distributions (ITT extrapolations) appeared reasonable. However, it was concluded that the Gompertz curve (ITT extrapolations) was more clinically reasonable in the short-term for UK patients. It was stated that the expected curve for the Int-2 and high-risk population may lie somewhere between these curves.

The clinicians were then shown the results of their preferred extrapolations for BAT and fedratinib on the same chart to confirm the relative impact of fedratinib on OS. It was advised that the fedratinib OS curve would not be expected to cross the BAT OS curve at any point; and it would be reasonable to prevent this from occurring in the economic model by assuming fedratinib OS would follow BAT OS in the long-term.

In the base case of the economic model, the Gompertz curve (Int-2/High-risk extrapolations) was selected as it closely aligned with clinicians' expectations for fedratinib survival (Figure 21). Clinical plausibility was prioritised over statistical fit, which indicated that the exponential distribution had the best fit to the observed data (see Table 50). In scenario analyses which include intermediate-1 patients, the more optimistic Weibull curve (ITT extrapolations) is used.

Figure 21: Parametric curves fit to overall survival data in the JAKARTA-2 intermediate-2 and high-risk population

Key: KM, Kaplan–Meier.

Table 50: Statistical fit of overall survival curves in JAKARTA-2

Distribution	AIC	BIC
Exponential	**232.61**	**235.01**
Generalised gamma	236.43	243.61
Gompertz †	234.58	239.37
Log-logistic	234.83	239.61
Log-normal	235.76	240.55
Weibull	234.59	239.38

Key: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
Notes: †, Selected distribution. Values in bold indicate best fitting parametric fit.

Figure 22 presents the base case survival curves used for both fedratinib and BAT treatment arms within the economic analysis. Based on the output of the pre-advisory board exercise (Table 48), it was not anticipated that curves would meet until around 15 years. However, it was acknowledged that long-term outcomes are highly uncertain. Parametric curves for BAT and fedratinib were initially selected sequentially by clinicians, and the selected curves crossed at 6 years, primarily due to the higher long-term plateaus in BAT extrapolations. However, clinical opinion was that the crossing of these curves would be implausible, as it was not expected that the survival of BAT patients would exceed that of fedratinib patients at any point. Therefore, upon the meeting of the curves, the long-term survival of fedratinib patients is set to equal that of the selected BAT curve.

Figure 22: A comparison of overall survival extrapolations between fedratinib and BAT

Key: BAT, best available therapy; KM, Kaplan–Meier; OS, overall survival.

The extrapolation of the alternative BAT KM sources presented in Figure 17 and used as options in the model are reported in Appendix L.5. An additional scenario modelling fedratinib OS extrapolation based on a surrogacy assumption between non-responders and responders was also explored as an option in the model. This scenario is described in Appendix L.7.

The time to death from transformation to AML and the re-estimation of overall survival based on this event is described in Appendix L.8.

B.3.3.4 Discontinuation

B.3.3.4.1 Time to treatment discontinuation

Time to treatment discontinuation (TTD) is estimated for patients on fedratinib. If a patient is still receiving fedratinib at the estimated TTD, they will discontinue treatment and move to BAT or palliative care. For patients on BAT, it is assumed the patient will remain on BAT until an alternative event occurs because no further treatments are available.

Therefore, for patients on fedratinib, there are three key factors which influence the explicit time of discontinuation:

1. 'Early discontinuation'. Early discontinuation refers to when a patient discontinues treatment before the response assessment. The proportion of early discontinuations was calculated from JAKARTA-2 trial data, where the patients who discontinued due to the clinical hold or death before EOC6 were excluded (see Table 51). The timing of the early discontinuation is estimated between 0 and 24 weeks using a uniform distribution. Early discontinuation and early death are mutually exclusive.

Table 51: Early discontinuation data

Parameter	Value	Source
Proportion of early discontinuations	[redacted]% (n = [redacted], N = [redacted])	JAKARTA-2 Int-2/high risk population. [redacted] patients discontinued before cycle 6 due to clinical hold (removed from n and N) and [redacted] patients died before EOC6 (removed from n)
	[redacted]% (n = [redacted], N = [redacted])	JAKARTA-2 ITT population PLD analysis. [redacted] patients discontinued before cycle 6 due to clinical hold (removed from n and N) and [redacted] patients died before EOC6 (removed from n)
Key: EOC6, end of cycle 6; n, number of early discontinuations; N, total patients; PLD, patient level data.		

2. ‘Non-response’. For non-responders, a parametric curve specific to non-responders from Week 24 is used to assign a TTD. In scenario analysis, a stopping rule is enabled, and non-responders discontinue immediately at Week 24.

For non-responder TTD, a Gompertz curve was chosen to reflect the expected limited time on treatment for non-responders in this population, despite not being the optimal statistical fit over the observed period (Table 52). Some of the other curves predicted long-term plateaus suggesting that non-responder patients would still be receiving fedratinib (if alive) beyond 10 years, which was not clinically appropriate (Figure 23). This choice of curve ensured that time-to-discontinuation was shorter on average for non-responders than responders. As the number of patients who were receiving fedratinib at 24 weeks but did not have spleen or symptom response (and did not have a censored TTD) was only 7, this was insufficient to estimate parametric curves for TTD. Therefore, the spleen non-responder TTD was used instead (number at risk = [redacted]).

Figure 23: Parametric curves fit to non-responder time-to-discontinuation data post-week 24 in JAKARTA-2 Int-2/High-risk patients

Key: KM, Kaplan–Meier.

Table 52: Statistical fit of non-responder time-to-discontinuation curves post-week 24 in JAKARTA-2 Int-2/High-risk patients

Distribution	AIC	BIC
Exponential	**54.452**	**55.343**
Generalised gamma	57.185	59.856
Gompertz †	56.419	58.200
Log-logistic	56.331	58.112
Log-normal	55.797	57.577
Weibull	56.387	58.167

Key: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
Notes: †, selected distribution.
Values in bold indicate best fitting parametric fit. Generalised Gamma did not converge so fit statistics could not be derived.

3. ‘Response’. A parametric curve specific to responders from Week 24 is used to assign a TTD.

For responder TTD, the exponential curve was chosen for its clinical plausibility, as other curves exhibited long-term plateaus (Figure 24). Statistical fit in the observed period for each parametric curve is presented in Table 53. Both the spleen and symptom responders were used to produce the parametric fit.

The time on fedratinib, overall, and split by responder status is presented in Appendix J.1.

Figure 24: Parametric curves fit to responder time-to-discontinuation data post-week 24 in JAKARTA-2 Int-2/High-risk patients

Key: KM, Kaplan–Meier

Table 53: Statistical fit of responder time-to-discontinuation curves post-week 24 in JAKARTA-2 Int-2/High-risk patients

Distribution	AIC	BIC
Exponential †	**89.680**	**91.317**
Generalised gamma	92.084	96.997
Gompertz	91.453	94.728
Log-logistic	91.392	94.667
Log-normal	90.660	93.935
Weibull	91.637	94.912

Key: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
Notes: †, selected distribution.
Values in bold indicate best fitting parametric fit.

B.3.3.4.2 Transition to palliative care

The palliative care health state reflects inpatient care in the final 8 weeks of life. At the time of fedratinib treatment discontinuation, if the patient's remaining life expectancy is ≤ 8 weeks, the patient will transition to palliative care, otherwise they will transition to BAT.

For patients in the BAT and AML health states, no explicit time to discontinuation is estimated. Therefore, a proportion is specified to determine how many patients will spend the final 8 weeks of life in palliative care (see Table 54). In the ruxolitinib NICE submission, 100% was chosen to reflect the end of life one-off cost applied to all patients. Due to the short time spent in this state relative to the time spent in the model, assumptions around the appropriate proportions to use for palliative care are likely to have little impact on the results.

Table 54: Proportions receiving palliative care from best available therapy and acute myeloid leukaemia health states

Parameter	Value	Source
Proportion receiving palliative care from BAT health state	100%	All patients in ruxolitinib NICE submission assigned end-of-life cost.
Proportion receiving palliative care from AML health state	100%	As above

Key: AML, acute myeloid leukaemia; BAT, best available therapy; NICE National Institute for Health and Care Excellence.

B.3.4. Measurement and valuation of health effects

B.3.4.1 Health-related quality-of-life data from clinical trials

The JAKARTA-2 trial evaluated health-related quality of life (HRQoL) using the Myelofibrosis Symptom Assessment Form Version 2.0 (MF-SAF V2.0) diary, Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF) questionnaire, the European Organisation for Research and Treatment of Cancer Quality of Life 30 Questionnaire Version 3.0 (EORTC QLQ-C30 V3.0), and the Patient Global Impression of Change (PGIC). No distinct preference-based measure was collected in the trial.

A preference-based index was identified for EORTC QLQ-C30 data when the data are split into eight dimensions (EORTC-8D).[95] Additionally, EORTC QLQ-C30 data and MF-SAF data had previously been combined to derive a preference-based index known as the MF-8D.[78]

Therefore, the MF-SAF V2.0 and EORTC QLQ-C30 V3.0 were analysed for the purpose of the model and mapped to derive utility values as described in the section below. The MF-SAF V2.0 was completed by patients daily through the first six cycles, via an electronic diary. The EORTC QLQ-C30 V3.0 was completed by patients on Day 1 of each treatment cycle up to Cycle 6, end of Cycle 6, Day 1 of Cycle 13, end of treatment, and at a 30-day follow-up visit.

B.3.4.2 Mapping

Generic preference-based measures of health, such as the EQ-5D, can be used to support the analysis of utility gains from treatments. In the absence of EQ-5D data, mapping algorithms are often used to link the outcomes from alternative measures of HRQoL to EQ-5D, or other generic preference-based measures.

There are some concerns regarding the ability of the generic EQ-5D to detect clinically meaningful changes in the HRQoL of patients with myelofibrosis.[78] This includes the exclusion of relevant symptoms such as nausea and vomiting.[78] Therefore, instead of mapping to the EQ-5D, two alternative methods were used to derive preference-based utility values from the JAKARTA-2 trial: (1) the MF-8D, and (2) the EORTC-8D.

The Myelofibrosis 8 dimensions (MF-8D) is a preference-based measure for MF which combines data from the MF-SAF and EORTC QLQ-C30 to generate utility scores. The MF-8D was the utility measure applied in the ruxolitinib NICE submission.[7]

Mukuria et al. (2015) developed the MF-8D as a condition-specific preference-based measure for MF to overcome the concerns related to using the generic EQ-5D and EORTC QLQ-C30 in the MF population.[78] Psychometric analyses of the performance of the EORTC QLQ-C30 against MF measures indicated that it does capture functioning and some generic symptom problems.[96] However, EORTC QLQ-C30 does not cover MF-specific symptoms (such as weight loss, itching, and night sweats) and is not as responsive as the MF-SAF over time.[97]

The patient population used to derive the scoring system for the MF-8D consisted of a clinical trial dataset of 309 patients from the Controlled Myelofibrosis Study with Oral JAK Inhibitor Treatment (COMFORT-I) trial.

The eight dimensions of the MF-8D are:

1. Physical functioning (from EORTC QLQ-C30)
2. Emotional functioning (from EORTC QLQ-C30)
3. Fatigue (from EORTC QLQ-C30)
4. Itchiness (from MF-SAF)
5. Pain under ribs on the left side (from MF-SAF)
6. Abdominal discomfort (from MF-SAF)
7. Bone or muscle pain (from MF-SAF)
8. Night sweats (from MF-SAF)

To calculate MF-8D, the closest data collection time point of MF-SAF was matched to each EORTC QLQ-C30 questionnaire. If an MF-SAF questionnaire could not be matched to within 2 weeks of the respective EORTC QLQ-C30 data collection date, then these measures were not used for the calculation of MF-8D. Figure 25 summarises MF-8D utility values by visit in JAKARTA-2 (minimum, lower quartile, median, upper quartile, maximum). These values suggest a pronounced increase in average health related quality of life for patients on fedratinib.

Figure 25: Utility values for MF-8D by visit in JAKARTA-2 Int-2/High-risk patients

Key: MF-8D, myelofibrosis 8 dimensions; n, sample size.

The second option explored was the EORTC 8 dimension (EORTC-8D), a preference-based measure for cancer which uses EORTC QLQ-C30 data to generate utility scores.[95]

The patient population used to derive EORTC-8D consisted of 655 patients with multiple myeloma in the VISTA trial.[95] As such, the EORTC-8D classification system was derived in a similar population to patients with myelofibrosis, but the lack of MF-specific data is a limitation of its use in this analysis.

The eight dimensions of the EORTC-8D are:

1. Physical functioning
2. Role functioning
3. Social functioning
4. Emotional functioning

5. Pain
6. Fatigue and sleep disturbance
7. Nausea
8. Constipation and diarrhoea

To calculate EORTC-8D the EORTC QLQ-C30 data from JAKARTA-2 were used. Figure 26 summarises EORTC-8D utility values by visit in JAKARTA-2. Consistent with MF-8D findings, EORTC-8D values show a pronounced increase in average health related quality of life for patients receiving fedratinib.

Figure 26: Utility values for EORTC-8D by visit in JAKARTA-2 Int-2/High-risk patients

Key: EORTC-8D, preference-based index from the EORTC QLQ-C30; n, sample size.

For application in the economic model, mixed effects models for both measures (MF-8D and EORTC-8D) were constructed to estimate utilities adjusted for covariates and for repeated measures within subjects, with the results presented in Table 55.

The mixed effects model was specified to account for potential prognostic factors. An intercept was specified, with covariates for baseline utility, sex and an indicator for response at the end of Cycle 6. The resultant utilities applied in the model are shown in Table 56 and Table 57. The MF-8D was selected to generate utility in the base case given that it was developed as a condition-specific measure and validated in myelofibrosis; whereas the lack of MF-specific data to inform the EORTC-8D is a limitation of its use in this analysis.

Utility analyses were also performed for the separate spleen and symptom response definitions and were included as options within the model. The results for these analyses are presented in Appendix L.

Table 55: Parsimonious mixed effects models for MF-8D and EORTC-8D for spleen or symptom response

Regression parameter	MF-8D			EORTC-8D		
	Coefficient	SE	p-value	Coefficient	SE	p-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline utility	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sex (Male)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen or symptom response (Y)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: EORTC-8D, European Organisation for Research and Treatment of Cancer 8 dimensions; MF-8D, myelofibrosis 8 dimensions; Ref, reference group, SE, standard error.
Note: Sex (Female) was used as the reference for the gender variable. Spleen or symptom response (N) was used as the reference for the spleen response variable

Table 56: MF-8D utilities to apply in the model

Utility	Implementation	Female	Male
Baseline	Baseline value	[redacted]	[redacted]
JAK response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
JAK non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
JAK loss response	Change from baseline, starting after loss of response	[redacted]	[redacted]
BAT response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT loss response	Change from baseline, starting after loss of response	[redacted]	[redacted]

Key: BAT, best available therapy; EORTC-8D; European Organisation for Research and Treatment of Cancer 8 dimensions; JAK, Janus kinase; MF-8D, myelofibrosis 8 dimensions.

Table 57: EORTC-8D utilities to apply in the model

Utility	Implementation	Female	Male
Baseline	Baseline value	[redacted]	[redacted]
JAK response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
JAK non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
JAK loss response	Change from baseline, starting after loss of response	[redacted]	[redacted]
BAT response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT loss response	Change from baseline, starting after loss of response	[redacted]	[redacted]

Key: BAT, best available therapy; EORTC-8D; European Organisation for Research and Treatment of Cancer 8 dimensions; JAK, Janus kinase; MF-8D, myelofibrosis 8 dimensions.

B.3.4.3 Health-related studies

The full details of the systematic searches conducted to identify relevant HRQoL data are outlined Appendix H. The SLR was supplemented by targeted searches to identify utility estimates specific to AML and palliative care health states.

B.3.4.4 Adverse reactions

Results of JAKARTA-2 demonstrated that fedratinib is generally well tolerated in patients with primary and secondary myelofibrosis.

In indirect treatment comparisons with the BAT arm of PERSIST-2 and SIMPLIFY-2 (see Table 58), the overall summary of safety in the fedratinib arm is comparable to the safety demonstrated in comparator trials, with a slightly higher incidence of adverse events (AEs) leading to treatment discontinuation. In both comparator studies, discontinuations were inconsistently reported in the BAT group because no-therapy was an acceptable BAT. This should be considered when interpreting the results presented in Table 58.

Table 58: Summary of treatment emergent adverse events reported for JAKARTA-2, PERSIST-2 (best available therapy arm only) and SIMPLIFY-2 (best available therapy arm only)

	JAKARTA-2 Int-2/High-risk: FEDR 400 mg (N=81)	PERSIST-2: BAT (N=98 [Safety population])	SIMPLIFY-2: BAT (N=52)
n (%) of subjects with at least one AE	[redacted]	87 (89)	46 (89)
n (%) of subjects with at least one Grade 3 or 4 AE	[redacted]	48 (49)	NR
n (%) of subjects with at least one SAE	[redacted]	30 (31)	12 (23)
n (%) of subjects who discontinued treatment due to AEs	[redacted]	4 (4)	1 (2)
n (%) of subjects with AEs leading to death	[redacted]	9 (9)*	4 (8)
n (%) of subjects with dose interruption for at least 7 consecutive days	[redacted]	10 (10)**	NR
n (%) of subjects with dose reduction	[redacted]	7 (7)	NR

Key: AE, adverse event; BAT, best available therapy; FEDR, fedratinib; SAE, serious adverse event.
Notes: *, percent is given for N=100; **, not specified whether the dose interruption was for a least 7 consecutive days.

In line with the approach taken in the ruxolitinib NICE submission, only non-haematological adverse events (AEs) grade ≥ 3 are explicitly modelled. This is because the impacts of thrombocytopenia, anaemia, and neutropenia (common haematological AEs in MF) on costs and utilities are assumed to be already captured by the model, in that:

- Costs of haematological AEs are counted in resource use estimates; and
- The impact on utilities of such AEs are assumed to be captured within the health state utility values.

Table 59 shows the observed rates of haematological AEs in JAKARTA-2.

Table 59: Frequency of grade ≥ 3 haematological adverse events

Adverse event	Fedratinib AEs (JAKARTA-2 [N=81])[43]
Anaemia	[redacted]
Thrombocytopenia	[redacted]
Leukopenia	[redacted]
Splenomegaly	[redacted]
Cytopenia	[redacted]
Febrile Neutropenia	[redacted]
Neutropenia	[redacted]
Thrombotic thrombocytopenic purpura	[redacted]
Key: AE, adverse events; N, total patients.	

The adverse events explicitly modelled for costs and disutility impacts were those included in the ruxolitinib NICE submission: abdominal pain, arthralgia, asthenia, back pain, bronchitis, cough, diarrhoea, dyspnoea, headache, nausea, peripheral oedema, pain in extremity, pyrexia and increased weight. Frequency data were identified in JAKARTA-2 for fedratinib (Table 60), from SIMPLIFY-2 for BAT (Table 61),[42] and from COMFORT-II for BAT post-fedratinib (Table 62) adjusted for average time of exposure in the model. The different sources for adverse events between the BAT arm and the BAT post-fedratinib consider how the proportion of therapies in BAT may influence AE proportions. As discussed in Section B.3.5.1, the BAT arm is assumed to have a BAT composition equal to that reported in SIMPLIFY-2[42], whereas BAT applied after fedratinib discontinuation is assumed not to consider any

JAK inhibitors. Therefore, the COMFORT-II BAT AE data is considered more representative of the BAT composition received by these patients. Although COMFORT-II AE data are for patients receiving first-line BAT, the absence of second-line AE data without JAK inhibitors made this the most appropriate data source.

In the absence of data on AE duration, all AEs were assumed to last for 4 weeks. The disutility values applied to AEs experienced in either treatment arm is reported in Table 63. The differing sources of AE event data for BAT had minimal difference on disutility between the BAT arms Table 63, and AE costs (See Section B.3.5.2).

Table 60: Frequency of grade ≥ 3 adverse events on fedratinib

Adverse event	n	N	Source
Abdominal pain	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Arthralgia	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Asthenia	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Back pain	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Bronchitis	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Cough	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Diarrhoea	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Dyspnoea	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Fatigue	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Headache	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Nausea	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Oedema peripheral	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Pain in extremity	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Pyrexia	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients
Weight increased	[redacted]	81	JAKARTA-2 PLD analysis of Int-2/High-risk patients

Key: CSR, clinical study report; n, number of patients with event; N, total number of patients.
Notes: Mean exposure to fedratinib in JAKARTA-2 was 0.539 years. Only adverse events with severity Grade ≥ 3 were considered.

Table 61: Frequency of grade ≥ 3 adverse events on best available therapy

Adverse event	n	N	Source
Abdominal pain	3	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Arthralgia	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2 (NR).
Asthenia	1	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Back pain	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2 (NR).
Bronchitis	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2 (NR).
Cough	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Diarrhoea	1	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Dyspnoea	1	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Fatigue	1	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Headache	1	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Nausea	1	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Oedema peripheral	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Pain in extremity	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2 (NR).
Pyrexia	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2.
Weight increased	0	52	Harrison et al., 2018, SIMPLIFY-2, Table 2 (NR).

Key: BAT, best available therapy; CSR, clinical study report; n, number of patients with event; N, total number of patients; NR, not reported.
Note: Mean exposure to fedratinib in SIMPLIFY-2 was 0.462 years. Only adverse events with severity grade ≥ 3 were considered.

Table 62: Frequency of grade ≥ 3 adverse events on best available therapy after fedratinib

Adverse event	n	N	Source
Abdominal pain	3	73	Cervantes et al. 2013, COMFORT-II, Table 2
Arthralgia	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Asthenia	1	73	Cervantes et al. 2013, COMFORT-II, Table 2
Back pain	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Bronchitis	1	73	Cervantes et al. 2013, COMFORT-II, Table 2
Cough	1	73	Cervantes et al. 2013, COMFORT-II, Table 2
Diarrhoea	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Dyspnoea	3	73	Cervantes et al. 2013, COMFORT-II, Table 2
Fatigue	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Headache	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Nausea	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Oedema peripheral	1	73	Cervantes et al. 2013, COMFORT-II, Table 2
Pain in extremity	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Pyrexia	0	73	Cervantes et al. 2013, COMFORT-II, Table 2
Weight increased	0	73	Cervantes et al. 2013, COMFORT-II, Table 2

Key: BAT, best available therapy; CSR, clinical study report; n, number of patients with event; N, total number of patients; NR, not reported.
Note: Only adverse events with severity grade ≥ 3 were considered.

Table 63: Disutility of included grade ≥ 3 adverse events

Adverse event	Disutility per event	Source
Abdominal pain	0.110	Tielemans et al. 2013[98], disutility for "gastrointestinal symptoms"
Arthralgia	0.220	Hollingworth et al. 2003[99], derived from a study on cancer-related back pain - utility is for pain/bone pain
Asthenia	0.090	Beusterien et al. 2010[100], disutility of grade 3-4 anaemia
Back pain	0.220	Assumed equal to arthralgia
Bronchitis	0.046	Assumed equal to cough
Cough	0.046	Doyle et al. 2008[101], disutility for cough in non-small-cell lung cancer population
Diarrhoea	0.047	Schremser et al. 2015[102], advanced lung adenocarcinoma patients
Dyspnoea	0.219	Lachaine et al. 2015[103], in relapsed acute promyelocytic leukaemia
Fatigue	0.073	Nafees et al. 2008[104], in non-small-cell lung cancer

Adverse event	Disutility per event	Source
Headache	0	No source identified
Nausea	0.048	Nafees et al. 2008[104], in non-small-cell lung cancer
Oedema peripheral	0	No source identified
Pain in extremity	0.105	Lachaine et al. 2015[103], disutility for pain
Pyrexia	0.110	Beusterien et al. 2010[100], disutility of grade 3-4 pyrexia
Weight increased	0	No source identified

Key: BAT, best available therapy; CSR, clinical study report; n, number of patients with event; N, total number of patients; NR, not reported.
Note: Mean exposure to fedratinib in SIMPLIFY-2 was 0.462 years. Only adverse events with severity grade ≥ 3 were considered.

B.3.4.5 Health-related quality-of-life data used in the cost-effectiveness analysis

Patients are assigned a baseline utility value in the model that is consistent between the intervention and the comparator. Health state utility values are then assigned as described below and in Table 64 to the following health states:

- Treatment health states
  - Response
  - Non-response
  - Loss of response
- AML
- Palliative care

Table 64: Summary of utility values for cost-effectiveness analysis

State	Assignment	Utility value: mean (standard error)	95% confidence interval	Reference in submission (section and page number)	Justification
Baseline utility	Baseline utility use for first 4 weeks after patient first receives treatment	Female: [redacted] ([redacted])	[redacted]	Section B.3.4 (page: 142 – Table 56)	Derived from JAKARTA-2 MF-8D analysis
		Male: [redacted]	[redacted]		

State	Assignment	Utility value: mean (standard error)	95% confidence interval	Reference in submission (section and page number)	Justification
Treatment: response	Change from baseline at 4 weeks if patient is classified as a responder	Female: [redacted][a] Male: [redacted][a]	[redacted]	Section B.3.4 (page: 142 – Table 56)	Derived from JAKARTA-2 MF-8D analysis
Treatment: non-response	Change from baseline at 4 weeks if patient is classified as a non-responder	Female: [redacted][a] Male: [redacted][a]	[redacted]	Section B.3.4 (page: 142 – Table 56)	Derived from JAKARTA-2 MF-8D analysis
Treatment: loss of response	Change from baseline if patients who are classified as responders lose response	Female: [redacted][a] Male: [redacted][a]	[redacted]	Section B.3.4 (page: 142 – Table 56)	Assumed to be the same as Treatment: non-response
AML	Utility value for patients who transition to AML health state	0.530 (0.053, [assumed 10% of mean])	0.426 – 0.633	Section B.3.4 (page: 149)	Derived from Pan et al. 2010[105]
Palliative care	Utility value for patients who transition to End of life health state who do not die	0.530 (0.053, [assumed 10% of mean])	0.426 – 0.633	Section B.3.4 (page: 150)	Capped at the value of the lowest utility (AML)

Key: AML, acute myeloid leukaemia; HS, health state; AR, adverse reaction.
Notes: [a] Multivariate normal distribution used to derive upper and lower bounds. Single SE value not available

In treatment health states, utility values depend on response status and are implemented as a change from baseline (CFB). The change in utility is assumed to start after 4 weeks of treatment, in line with the assumptions applied in the ruxolitinib NICE submission.

The utility value for AML was identified in a systematic review of health state utility values for AML.[106] A wide range of AML utilities were reported, as different groups of patients with AML were included in the study: patients undergoing induction treatment (range 0.524 to 0.67); patients in relapse (range 0.50 to 0.53); patients in remission post-chemotherapy (range 0.81 to 0.91); and patients post-stem cell

transplant (range 0.71 to 0.83).[106] The most appropriate utility value in this context was for patients with ‘secondary AML’ (0.53) as patients in the model have progressed to AML from myelofibrosis.[105]

The utility value for palliative care was estimated using the EQ-5D in patients with either end-stage breast, prostate, or colorectal cancer (0.59).[107] It was not possible to identify a palliative care utility value specific to MF, since published quantitative data on utility in patients approaching the end of life in palliative care are rare.[108] An unexpected result is that the utility value identified for palliative care (0.59) is greater than that of secondary AML (0.53). This may truly reflect patient HRQoL. However, the discrepancy may be due to a selection bias, in that patients who are asked to complete patient reported outcome measures in palliative care may be a ‘healthier’ subset of the palliative care population. Alternatively, the discrepancy may be due to palliative care data not being specific to MF, or due to limitations of the EQ-5D which showed a pronounced ceiling effect (13% of patients reported full health).[107] The effect of this parameter is low due to the limited number of weeks spent in the palliative care health state. The model can cap palliative care utility by other health state utilities.

As an alternative scenario, utilities from the ruxolitinib SMC submission can be used (see Table 65). Most utilities in the ruxolitinib NICE submission were redacted. In ruxolitinib submissions, a supportive care health state was included which was associated with a decrement in utility every 24 weeks. Based on clinician feedback that 'supportive care' is equivalent to BAT, only a BAT health state is included in the fedratinib model, in which the option exists to replicate the worsening utility approach. Palliative care was not modelled as a health state in the ruxolitinib submissions, so for this scenario a utility value of 0.59 (derived from Färkkilä et al. 2014[107]) capped at the lowest utility value was used for patients transitioning to palliative care.

Table 65: Summary of utility values applied in ruxolitinib modelling (TA386)

State	Assignment	Utility value: mean (standard error)	95% confidence interval	Justification
Baseline utility	Baseline utility use for first 4 weeks after patient first receives treatment	0.732 (0.073, [assumed 10% of mean])	0.577 – 0.862	Taken from Ruxolitinib SMC DAD[65]
JAK Treatment: response	Change from baseline at 4 weeks if patient receiving JAKi is classified as a responder	0.153 (0.015, [assumed 10% of mean])	0.124 – 0.184	Taken from Ruxolitinib SMC DAD[65]
JAK Treatment: non-response	Change from baseline at 4 weeks if patient receiving JAKi is classified as a non-responder	0.037 (0.004, [assumed 10% of mean])	0.030 – 0.045	Taken from Ruxolitinib SMC DAD[65]
BAT	Change from baseline	0	0	Taken from Ruxolitinib SMC DAD[65] No response was allowed for BAT patients in model
Worsening utility	Utility of patients receiving BAT is reduced every 24 weeks by this utility decrement.	0.025 (0.003, [assumed 10% of mean])	0.020 – 0.030	Taken from Ruxolitinib SMC DAD[65]
AML	Decrement applied to patient utility upon transitioning to AML	0.257 (0.026, [assumed 10% of mean])	0.208 – 0.309	Taken from Ruxolitinib NICE submission TA386[7]

Key: AML, acute myeloid leukaemia; BAT, best available therapy; DAD, detailed advice document; JAK, Janus Kinase; JAKi, JAK inhibitor

The proportion of patients experiencing each adverse event was informed by JAKARTA-2, SIMPLIFY-2 and COMFORT-II data (See Section B.3.4.4) accounting for mean time of treatment exposure. These data were used to inform annual disutility values, which were multiplied by the years spent on treatment in the model. The values used are presented in Table 66.

Table 66: Annual adverse event disutilities

Starting treatment:	Annual disutility
Fedratinib	0.001
BAT	0.003
BAT, after two JAK inhibitors	0.003
Key: BAT, best available therapy; JAK, Janus Kinase	

To account for the natural decline in quality of life over time, utilities were adjusted throughout the simulation based on the patient's age. The adjustment is based on a formula published by Ara and Brazier (2010).[109] Ara and Brazier (2010) used data from a large sample of the UK general population (n = 26,679) to fit a regression to predict mean health state utility values based on age and gender.[109] The formula is presented in Equation 1.

Equation 1: Calculation of general population utility scores, to inform age-related utility adjustment in the model

$$\text{Utility} = 0.9508566 + (0.0212126 * \text{male}) - (0.0002587 * \text{age}) - (0.0000332 * \text{age}^2)$$

B.3.5. Cost and healthcare resource use identification, measurement and valuation

An SLR was conducted to identify any relevant cost and healthcare resource use data associated with the treatment of patients with myelofibrosis. Appendix I outlines the methods used in the SLR. The cost and healthcare resource use applied in the model were primarily based on standard national tariffs and resource use data presented in NICE TA386.

B.3.5.1 Intervention and comparators' costs and resource use

Annual acquisition and administration costs for BAT are calculated based on the proportions specified for BAT composition. The available sources reporting the composition of BAT arm are presented in Table 67, with the reported composition of BAT for each source reported in Table 68. Of the data available, SIMPLIFY-2 was identified as the most appropriate source for the composition of BAT, this was because of the following reasons:

- The ruxolitinib proportion and overall BAT composition used in SIMPLIFY-2 was identified as the most realistic values for clinical practice by a UK advisory board[18]
- PERSIST-2 included patients which had not necessarily received ruxolitinib and only included patients with a platelet count < 100 x 10^9/L, and therefore was less comparable to the JAKARTA-2 study population than SIMPLIFY-2
- Schain et al.[15] presented results exclusively from Sweden and Norway, which clinicians decided may be inappropriate for a UK setting[18]
- The HMRN 2020 report[12] covering the region Yorkshire and the Humber and Yorkshire Coast Cancer networks were captured and only [redacted] observations of treatments were reported following ruxolitinib discontinuation. The sample size is therefore too small to produce reliable values

Table 67: Studies reporting BAT composition arm in myelofibrosis

Study	Study details
SIMPLIFY-2[42]	Phase III multicentre randomised open-label clinical trial with 52 patients in the BAT arm. Patients were currently or previously treated with ruxolitinib but did include intermediate-1 patients.
PERSIST-2[41]	Phase III multicentre randomised open-label clinical trial with 72 patients in the BAT arm. Patients could receive up to 2 JAK2 inhibitors. Includes intermediate-1 patients. Only included patients with a platelet count < 100 x 10^9/L
Schain 2019[15]	Study was a retrospective analysis in patients from Sweden and Norway
HMRN 2020[12]	Resource utilisation and outcomes of patients in Yorkshire and the Humber and Yorkshire Coast Cancer networks. Only [redacted] observations of subsequent treatments following ruxolitinib were captured in the HMRN report, so only these patients were used to calculate BAT proportion in second line.
Key: BAT, best available therapy; HMRN, Haematological Malignancy Research Network.	

Table 68: BAT composition observed in different sources

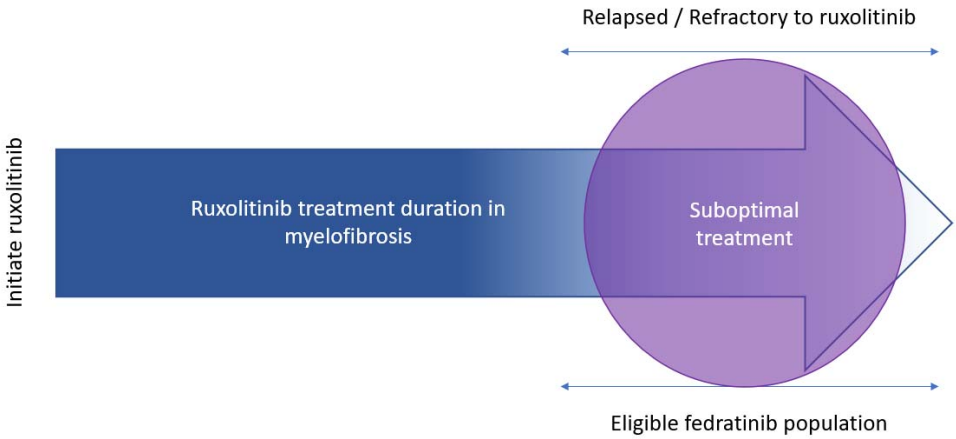
Treatment	SIMPLIFY-2 (n=52)	PERSIST-2 (n=98)	Schain 2019 (n=37)	HMRN 2020 (n=[redacted])
Anagrelide	0% (NR)	0.0%	0.0%	[redacted]
Busulfan	0% (NR)	0.0%	8.1%	[redacted]
Cytarabine	0% (NR)	0.0%	0.0%	[redacted]
Danazol	0% (NR)	5.1%	5.4%	[redacted]
Decitabine	0% (NR)	2.0%	0.0%	[redacted]
Hydroxycarbamide (hydroxyurea)	23.1%	19.4%	32.4%	[redacted]
Interferon alfa	0% (NR)	1.0%	0.0%	[redacted]
Peginterferon alfa-2a	0% (NR)	1.0%	5.4%	[redacted]
Prednisolone	5.8%	6.6%	32.4%	[redacted]
Prednisone	5.8%	6.6%	32.4%	[redacted]
Thalidomide	0% (NR)	3.1%	5.4%	[redacted]
Ruxolitinib	88.5%	44.9%	0.0%	[redacted]
Actively treated	100.0% (NR)	80.6%	100.0% (NR)	[redacted]
Key: BAT, best available therapy; HMRN, Haematological Malignancy Research Network; NR, not reported.				

For patients discontinuing fedratinib in the model, clinical opinion was that there would be a likely reduction in patients receiving JAK-inhibitors after previously failing

both ruxolitinib and fedratinib. In the absence of data informing third-line therapies after failure of two JAK-inhibitors, the proportions of therapies in the BAT arm were re-weighted to remove ruxolitinib (Table 69).

As noted in Section B.3.2, clinical experts advised that patients are often continued on ruxolitinib while achieving a suboptimal response, as no other targeted therapeutic options are available (Figure 27).

Figure 27: Schematic representation of the current treatment duration in those that respond to ruxolitinib



Key: BAT, best available therapy

If fedratinib was available, there would be an opportunity to switch patients who are relapsed, refractory or intolerant to ruxolitinib onto an effective therapy. In current practice, if the composition for BAT were defined from the point at which patients *should* discontinue ruxolitinib, the UK clinical experts indicated they would expect ruxolitinib use in this population to be similar or greater than that used in SIMPLIFY-2 (89%).[18] Therefore, this proportion of ruxolitinib use is applied in the BAT composition, and scenarios are presented in the model which assess lower proportions of ruxolitinib use within BAT. The assumptions surrounding the proportion of ruxolitinib use within BAT and its wider implications on model inputs and calculations are detailed in Section B.3.6.1.

Table 69: Best available therapy composition

Treatment	Active input, BAT after ruxolitinib (%)[a]	Active input, BAT after ruxolitinib and fedratinib (%)[b]
Anagrelide	0.0%	0.0%
Busulfan	0.0%	0.0%
Cytarabine	0.0%	0.0%
Danazol	0.0%	0.0%
Decitabine	0.0%	0.0%
Hydroxycarbamide (hydroxyurea)	23.1%	66.7%
Interferon alfa	0.0%	0.0%
Peginterferon alfa-2a	0.0%	0.0%
Prednisolone	5.8%	16.7%
Prednisone	5.8%	16.7%
Thalidomide	0.0%	0.0%
Ruxolitinib	88.5%	0.0%

Key: BAT, best available therapy.
Notes:
[a]Proportions taken from SIMPLIFY-2[42]
[b]Proportions taken from SIMPLIFY-2[42], ruxolitinib was then set to 0%, and the remaining treatments were reweighted to maintain the proportion actively treated.

Drug acquisition costs were sourced primarily from the Monthly Index of Medical Specialities (MIMS) online database.[110] For drugs available in generic form, acquisition costs were sourced from the Drugs and pharmaceutical electronic Market Information Tool (eMIT), because eMIT costs are based on actual purchases made by the NHS, as opposed to list prices.[111]

Where multiple costs were identified for treatments in BAT, the cost was selected based on the lowest cost per milligram, so long as the strength was a valid option for the dose.

For ruxolitinib use, an additional 5% wastage assumption was applied, in line with preferred ERG assumptions from TA386.[7] This attempts to account for frequent dose adjustments on ruxolitinib, which results in the remaining tablets within a pack being discarded. In contrast, the unit dose of fedratinib is 100 mg per tablet and dose adjustments on fedratinib are implemented in increments of 100 mg (e.g. a patient may move from 400 mg daily, to 300 mg, to 200 mg when they experience adverse

events). Overall, fedratinib wastage was applied by costing per pack in line with other treatments in the model.

Averages for patient weight and body surface area (BSA) were used to calculate doses where appropriate. Acquisition costs for oral therapies are presented in Table 70. Acquisition costs for intravenous therapies are presented in Table 71.

Table 70: Drug acquisition unit costs (oral therapies)

Treatment	Pack size	Unit size	Unit type	Pack cost	Cost per unit	Reference
Fedratinib	120	100 mg	Tablet	[redacted]	[redacted]	Net price provided by Celgene, a BMS company
Ruxolitinib	56	5 mg	Tablet	**£1,428**	£25.50	MIMS[112]
	56	10 mg	Tablet	**£2,856**	£51.00	
	56	15 mg	Tablet	**£2,856**	£51.00	
	56	20 mg	Tablet	**£2,856**	£51.00	
Busulfan	25	2 mg	Tablet	**£69.02**	£1.3804	NHS Drug Tariff[113]
Danazol	60	100 mg	Tablet	**£10.07**	£0.0017	MIMS[114]
	60	200 mg	Tablet	£36.32	£0.0030	
Hydroxycarbamide (hydroxyurea)	100	500 mg	Tablet	**£9.56**	£0.0002	eMIT[111]
Prednisolone	28	1 mg	Tablet	£0.18	£0.0064	eMIT[111]
	28	2.5 mg	Tablet (gastro resistant)	£0.61	£0.0087	
	28	2.5 mg	Tablet	£0.55	£0.0079	
	30	20 mg	Tablet	£3.77	£0.0063	
	56	25 mg	Tablet	£19.23	£0.0137	
	28	5 mg	Tablet (gastro resistant)	£0.63	£0.0045	
	30	5 mg	Tablet (soluble)	£14.89	£0.0993	
	28	5 mg	Tablet	**£0.31**	£0.0022	
Prednisone	30	1 mg	Tablet	£26.70	£0.8900	BNF[79]
	30	2 mg	Tablet	£26.70	£0.4450	

Treatment	Pack size	Unit size	Unit type	Pack cost	Cost per unit	Reference
	30	5 mg	Tablet	**£26.70**	£0.1780	
Thalidomide	28	50 mg	Capsule	**£298.48**	£0.2132	MIMS[115]

Key: BNF, British National Formulary; eMIT, electronic market information tool; mg, milligrams; MIMS, Monthly Index of Medical Specialities; NHS, National Health Service.
Notes: For each treatment, the cost was selected on the basis of the lowest cost per mg while being a valid option for the dose. The selected options used within the model are indicated in bold.

Table 71: Drug acquisition unit costs (intravenous therapies)

Treatment	Pack size	Unit size	Unit type	Pack cost	Cost per unit	Reference
Cytarabine	5	5 x 5 ml 20 mg/ml	Solution for vial	£20.48	£0.04	MIMS[116]
	5	5 x 1 ml 100 mg/ml	Solution for vial	£26.93	£0.05	
	1	1 x 10 ml 100 mg/ml	Solution for vial	**£37.05**	£0.04	
Decitabine	1	50 mg	Powder	**£970.86**	£19.42	MIMS[117]
Interferon- alfa	1	3 million IU	Pre-filled syringe	**£14.20**	£4.73	MIMS[118]
	1	4.5 million IU	Pre-filled syringe	£21.29	£4.73	
	1	6 million IU	Pre-filled syringe	£28.37	£4.73	
Peginterferon alfa-2a	1	90 microgram/0.5 ml	Pre-filled syringe	£76.51	£0.85	MIMS[119]
	1	135 microgram/0.5 ml	Pre-filled syringe	£107.76	£0.80	
	4	180 microgram/0.5 ml	Pre-filled syringe	**£497.60**	£0.69	

Key: mg, milligrams; MIMS, Monthly Index of Medical Specialties.
Notes: For each treatment, the cost was selected on the basis of the lowest cost per mg while being a valid option for the dose. The selected options used within the model are indicated in bold.

Oral treatments (such as JAK-inhibitors) are assumed to have no associated administration costs. Self-administered treatments (peginterferon alfa-2a) are assumed to have no administration cost to the healthcare payer. Treatments administered by injection are assigned a flat cost per administration taken from NHS Reference Costs (Table 72).[120] The DES enables acquisition and administration costs to be accumulated at the point of prescription and administration. This ensures that wastage due to death or discontinuation is included.

Table 72: Drug administration unit costs

Method of administration	Cost	Notes on costing	Reference
Injection	£332	Cost applied per administration	NHS Reference Costs.[120] Code: SB15Z
Oral	£0	Cost applied per prescription	Assumption
Self-administration	£0	Cost applied per administration	Assumption
Key: NHS, National Health Service.			

As health care resource use data are typically skewed, with outliers and a large proportion of patients having no reported resource utilisation, a large sample size is required to derive adequate estimates of treatment-specific resource utilisation.[7]

The JAKARTA-2 study had a relatively small sample size for resource use calculations, with short follow-up; therefore, resource use in the model is primarily informed by the ruxolitinib NICE submission, which leveraged data from three sources:

- HMRN audit (2016):[121] UK audit of clinical management, resource utilisation and outcome in primary and secondary myelofibrosis
- The ROBUST study:[122] a phase II study that was done in the UK (n=48). It included patients with intermediate-1, intermediate-2 and high-risk disease
- The JUMP study:[123] A phase III expanded-access trial designed to assess the safety and efficacy of ruxolitinib in patients with high-risk, intermediate-2 risk or intermediate-1 risk disease. This study did not include any patients from the UK.

The HMRN audit in 2016 and the ROBUST study were UK-specific studies, [121, 122] and were used to inform resource use for patients receiving BAT. The ruxolitinib NICE submission used either assumptions or the JUMP study to inform the change in resource use associated with ruxolitinib, relative to BAT.

The HMRN audit in 2016 assessed a time-period whilst ruxolitinib was approved in the Cancer Drugs Fund by NICE. Where possible, inputs for this submission were updated using the HMRN 2020 audit. The updated HMRN audit also included resource use for patients who received ruxolitinib, so this was used to recalculate the relative impact of a JAK inhibitor on resource use over time. The base case assumptions for the model apply the most up-to-date data. However, the original values used in TA386 were added as an option within the model.

Table 73: Weekly resource use on best available therapy – NICE TA386

Resource	Best available therapy	Source
A&E visit	0.013	ROBUST - NICE 2016, TA386, committee papers (ACD), Table 47[7, 122]
FBC & U&E	0.32	HMRN Audit - NICE 2016, TA386, committee papers (ACD), Table 47[7, 121]
Hospital night	0.15	HMRN Audit - NICE 2016, TA386, committee papers (ACD), Table 47[7, 121]
Outpatient visit	0.22	HMRN Audit - NICE 2016, TA386, committee papers (ACD), Table 47[7, 121]
Primary care visit	0.03	ROBUST - NICE 2016, TA386, committee papers (ACD), Table 47[7, 122]
RBC unit transfusion	0.16	Assumption - NICE 2016, TA386, committee papers (ACD), Table 47[7]
Urgent care	0.003	ROBUST - NICE 2016, TA386, committee papers (ACD), Table 47[7, 122]

Key: A&E, Accident & Emergency; ACD, appraisal consultation document; FBC, full blood count; HMRN, Haematological Malignancy Research Network; NICE, National Institute for Health and Care Excellence; RBC, red blood cell; TA, technology appraisal; U&E, urea & electrolytes.

Table 74: Weekly resource use– HMRN 2020

Resource	All patients	Patients treated with ruxolitinib
A&E visit	[redacted]	[redacted]
FBC & U&E	NR	NR
Hospital night	[redacted]	[redacted]
Outpatient visit	[redacted]	[redacted]
Primary care visit	NR	NR
RBC unit transfusion	NR	NR
Urgent care	[redacted]	[redacted]

Key: A&E, Accident & Emergency; ACD, appraisal consultation document; FBC, full blood count; HMRN, Haematological Malignancy Research Network; NICE, National Institute for Health and Care Excellence; RBC, red blood cell; TA, technology appraisal; U&E, urea & electrolytes.

Table 75: Resource use on JAK-inhibitor relative to best available therapy – TA386, updated using available HMRN 2020 values

Resource	Up to Week 12	Up to Week 24	Up to Week 36	Up to Week 48	Up to Week 108	Up to Week 144	Beyond Week 144	Source
A&E visit	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	HMRN 2020[12]
FBC & U&E	+4.00%	-82.60%	-82.60%	-82.60%	-82.60%	-82.60%	-82.60%	Assumptions - NICE 2016, TA386, committee papers (ACD), Table 47[7]
Hospital night	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	HMRN 2020[12]
Outpatient visit	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	HMRN 2020[12]
Primary care visit	0.00%	-36.70%	-58.20%	-81.70%	-97.70%	-97.70%	-97.70%	JUMP - NICE 2016, TA386, committee papers (ACD), Table 47[7, 124]
RBC unit transfusion	+43.30%	+43.30%	+10.00%	+10.00%	+10.00%	-23.30%	-58.30%	Assumptions - NICE 2016, TA386, committee papers (ACD), Table 47[7]
Urgent care	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	HMRN 2020[12]

Key: A&E, Accident & Emergency; ACD, appraisal consultation document; FBC, full blood count; HMRN, Haematological Malignancy Research Network; NICE, National Institute for Health and Care Excellence; RBC, red blood cell; TA, technology appraisal; U&E, urea and electrolytes.

Table 76: Unit costs of monitoring and resource use

Resource	Sourced unit cost	Price year	Cost per event (adjusted to 2019)[125, 126]	Source
A&E visit	£166.05	2019	£166.05	NHS Reference Costs 2018/19 (Accident & Emergency)[120]
FBC & U&E	£70.00	2019	£70.00	Private Patient Tariff 2019 (Dorset County Hospital, Full blood count and U&E profile)[127]
Hospital night	£589.07	2019	£589.07	NHS Reference Costs 2018/19 (Non Elective Inpatients Excess Bed Day)[120]
Outpatient visit	£166.51	2019	£166.51	NHS Reference Costs 2018/19 (WF01A - Clinical Haematology, Non-Admitted Face-to-Face)[120]
Primary care visit	£29.00	2019	£29.00	PSSRU Unit Costs 2019 (GP consultation)[125]
RBC unit transfusion	£235.00	2001	£371.70	Varney 2003 (cost per RBC unit)[128]
Urgent care	£153.86	2019	£153.86	PSSRU Unit Costs 2019 (Acute medical unit)[120]

Key: A&E, accident and emergency; FBC, full blood count; NHS, National Health Service; PSSRU, Personal Social Services Research Unit; RBC, red blood cell; U&E, urea and electrolytes.

The overall resource use was calculated according to the active proportion of JAK-inhibitor in each of the treatment arms (100% for fedratinib before discontinuation; 88.5% for BAT as a comparator; 0% for BAT after fedratinib discontinuation) and is presented in Table 77.

The calculation of the costs of the individual treatment arms could be interpreted as conservative given the large difference in the proportion of patients expected to respond between the treatment arms. It could be further argued that this approach disadvantages fedratinib because 'BAT after fedratinib' is indicative of a loss of response in the fedratinib arm, however the same loss of response in the BAT arm is not associated with higher resource costs in the model. Because of the limitations of the data source being split by JAK-inhibitor administration and not able to consider a relationship between response and resource use, the option is available in the model to set 'BAT as comparator' and 'BAT after fedratinib' costs as equal to 'Fedratinib'.

Table 77: Resource use by period for fedratinib and best available therapy.

Cost per week	Fedratinib	BAT as comparator	BAT after fedratinib
Cost per week: 0 - 12 weeks	£210.24	£210.29	£210.66
Cost per week: 12 - 24 weeks	£190.41	£192.75	£210.66
Cost per week: 24 - 36 weeks	£170.35	£175.00	£210.66
Cost per week: 36 - 48 weeks	£170.08	£174.76	£210.66
Cost per week: 48 - 108 weeks	£169.89	£174.59	£210.66
Cost per week: 108 - 144 weeks	£150.09	£157.08	£210.66
Cost per week: 144+ weeks	£129.27	£138.66	£210.66

An additional separate resource use consideration applied exclusively to the fedratinib treatment arm is for thiamine testing and supplementation. Thiamine testing is anticipated to occur at baseline, then once every month for the first 3 months, then once every 3 months.[129] Clinical input indicated that thiamine testing would be conducted alongside other routine tests, and therefore no extra hospital visits would be required or costed. However, it was additionally advised that few centres in the UK have the capacity to conduct thiamine tests, therefore, the samples are sent to centres which do. Therefore, a provider-to-provider cost of £31 per test was identified and applied to all test instances in the model.[130]

Within the JAKARTA-2 CSR, of the 28 patients that were tested for thiamine deficiency upon discontinuing fedratinib, 3 were found to have thiamine levels below normal. As such, 10.71% was used as the input for patients requiring thiamine supplementation. For simplicity in the economic model, it is assumed that all patients requiring thiamine supplementation incur the cost of a full 90-day course, which is costed once upon the initiation of fedratinib and once again upon discontinuation. Thiamine dose may vary between 50mg – 300mg per day according to the severity of the deficiency; 200mg per day was assumed as an average dose per patient. The unit cost of thiamine is presented in Table 78.

Table 78: Thiamine unit costs

Treatment	Pack size	Unit size	Unit type	Pack cost	Cost per unit	Reference
Thiamine	100	50	Tablet	4.35	0.04	MIMS[131]
	100	100	Tablet	**5.83**	0.06	
Notes: The selected option used within the model is indicated in bold.						

B.3.5.2 Adverse reaction unit costs and resource use

Adverse event costs were identified from National Health Service (NHS) Reference Costs,[120] the Personal Social Services Research Unit (PSSRU) Unit Costs of Health and Social Care,[125] and other sources accessed in the ruxolitinib NICE submission.[7] Where NHS Reference Costs were used, weighted averages of relevant currency/service codes were calculated. Sources were consistent with those selected for the ruxolitinib NICE submission, with values taken from updated publications where available. Unit costs for AEs are presented in Table 79. The unit costs for AEs were combined with the AE frequency data reported in Section B.3.4.4 to produce the annual AE costs in Table 80.

Table 79: Unit costs of adverse events in the economic model

Adverse event	Sourced cost per event	Price year of cost source	Cost per event (adjusted to 2019)[125, 126]	Source (cost)
Abdominal pain	£634.50	2019	£634.50	NHS Reference Costs 2018-19 (weighted average: FD05A, FD05B)[120]
Arthralgia	£157.20	2019	£157.20	NHS Reference Costs 2018-19 (service code: 19)[120]
Asthenia	£12.00	2014	£12.84	NICE 2014, TA316, evaluation report 4, Table 68[132]
Back pain	£808.94	2019	£808.94	NHS Reference Costs 2018-19 (weighted average: HC32G-HC32K)[120]
Bronchitis	£40.45	2019	£40.45	PSSRU 2019 (GP consultation)[125] & MIMS 2020 (course of clarithromycin)[133]
Cough	£40.45	2019	£40.45	PSSRU 2019 (GP consultation)[125] & MIMS 2020 (course of clarithromycin)[133]
Diarrhoea	£39.99	2019	£39.99	PSSRU 2019 (GP consultation)[125] & MIMS 2020 (course of loperamide)[134]
Dyspnoea	£0.00	2014	£0.00	NICE 2016, TA386, committee papers (ACD), Table 46[7]
Fatigue	£12.00	2014	£12.84	NICE 2014, TA316, evaluation report 4, Table 68[132]
Headache	£117.00	2004	£161.76	McCrone et al., J Headache Pain 2011;12:617–23[135]
Nausea	£39.99	2019	£39.99	PSSRU 2019 (GP consultation)[125] & MIMS 2020 (course of ondansetron) [136]
Oedema peripheral	£914.00	2014	£978.12	NICE 2014, TA316, evaluation report 4, Table 68[132]
Pain in extremity	£157.20	2019	£157.20	NHS Reference Costs 2018-19 (service code: 191, Pain Management)[120]
Pyrexia	£3,076.99	2009	£3,581.83	Woods et al., Value Health 2012;15:759–70.[137]
Weight increased	£78.00	2019	£78.00	PSSRU 2019 (2 GP consultations)[125]

Key: ACD, appraisal consultation document; GP, general practitioner; MIMS, Monthly Index of Medical Specialities; NICE, National Institute for Health and Care Excellence; PSSRU, Personal Social Services Research Unit; TA, technology appraisal.

Table 80: Annual AE costs

Treatment	Annual cost
Fedratinib	£27.09
BAT as comparator	£98.83
BAT after fedratinib	£88.79

B.3.6. Summary of base-case analysis inputs and assumptions

B.3.6.1 Proportion of ruxolitinib within BAT

A significant limitation of the available data was that there was no head-to-head comparison of fedratinib versus BAT in the appropriate indication. Therefore, assumptions were required to inform the composition of BAT, proportion of patients with response, estimate overall survival, and other inputs. Searches identified relevant studies reporting response (Table 31), survival (Figure 17) and BAT composition (Table 67) to populate the model with values best representing the decision problem. It was found that a primary driver of costs outcomes in the model was the proportion of ruxolitinib within the BAT treatment arm, therefore the input for this value use in the base case was carefully considered using available evidence and clinical opinion.

As discussed in Section B.3.5.1, of the identified studies that specified the proportion of therapies within the BAT treatment arm, the BAT arm reported in SIMPLIFY-2 was considered to be the most appropriate. The SIMPLIFY-2 study reported that 88.5% patients who failed ruxolitinib were nevertheless treated with ruxolitinib as part of best available therapy. This study was considered the most appropriate because of the overlap between the SIMPLIFY-2 and JAKARTA-2 study populations, the number of patients observed, and clinical opinion that patients are often continued on ruxolitinib while achieving a suboptimal response as no other targeted therapeutic options are available.[18]

Of the studies reporting BAT composition in myelofibrosis (Table 67), only SIMPLIFY-2 and PERSIST-2 were feasible to include in ITC analyses to determine the proportion of responders (See section B.3.3.2). In contrast to NICE TA386, this submission allows patients receiving BAT to respond and therefore experience

improvements in HRQoL. This approach ensures that ITC and economic model outcomes are suitably representative of the composition of BAT used in each of the data sources. The base case response ITC uses SIMPLIFY-2 data since only naïve comparisons could be performed with the available PERSIST-2 data (See section B.3.3.2). Therefore, the response outcomes used in the model are representative of a BAT arm in which 88.5% patients have received ruxolitinib.

SIMPLIFY-2 was also used to inform BAT adverse event frequency in the model, such that adverse events are based on there being 88.5% ruxolitinib in the BAT arm. This in turn influences the costs of adverse events, and the utility associated with adverse events.

Where appropriate and feasible, other costs such as drug and resource use costs, are weighted in the model by the proportion of ruxolitinib use. However, HRQoL values in the treatment states are dependent on whether a patient responds. Given that only summary data were available from external studies, it was not possible to separate out the proportion of patients in BAT responding on ruxolitinib and responding, not on ruxolitinib. Therefore, response data is inflexible to changes in BAT composition.

As mentioned above, clinical experts advised that patients are often continued on ruxolitinib while achieving a suboptimal response[18]; however, limited OS data was identified for such patients continuing ruxolitinib[138]. Therefore, it was assumed that OS for suboptimal treatment with ruxolitinib would be comparable to BAT OS, and therefore BAT OS in the model is independent of the proportion of ruxolitinib in BAT. The SIMPLIFY-2 BAT OS was unsuitable for informing the model BAT OS owing to crossover with momelotinib at week 24 of the study. However, it was reported that at 24 weeks 21% patients had died in the BAT arm,[138] which appears consistent with the available BAT KM data presented in Figure 17.

A summary of how the proportion of ruxolitinib in BAT is used in the model is presented in Table 81. This table shows that some inputs in the model are intrinsically linked to a high proportion of ruxolitinib use through the SIMPLIFY-2 study (e.g. response, HRQoL). Furthermore, the use of a high proportion of ruxolitinib in BAT is supported by clinical opinion[18] and published literature.[40, 80, 81] It

is likely that lower proportions of ruxolitinib would not be appropriate owing to the inconsistency of assumptions between the model and the data sources, as well as being potentially not representative of UK clinical practice.[18]

Table 81: How the proportion of ruxolitinib in BAT is used in the model

Model outcome	Flexible to changes in ruxolitinib %?	Rationale
Response	No	The data informing BAT response did not provide a breakdown of results by individual treatment. Therefore, ITC results are subject to the compositions used in the original data (either SIMPLIFY-2 or PERSIST-2).
HRQoL	No	Treatment health-state utility is driven by response, which is directed by the response ITC. As described above, response is not influenced by the BAT composition applied in the model. Adverse events are informed by SIMPLIFY-2; therefore, disutility is representative of 88.5% ruxolitinib in BAT
Costs: Drugs	Yes	The overall BAT drug costs are derived by weighting individual drug costs according to proportion of the drug included in BAT.
Costs: Resource use	Yes	Resource use was sourced primarily from TA386 which reported BAT and ruxolitinib (JAKi) resource use. Costs are weighted by JAKi proportion.
Cost: Adverse events	No	Adverse events are informed by SIMPLIFY-2. Therefore, adverse event costs are representative of 88.5% ruxolitinib in BAT.
Survival	No	Limited data was available for ruxolitinib versus non-ruxolitinib survival after ruxolitinib treatment failure. The proportion of ruxolitinib in BAT was not expected to have a significant impact on overall survival; this assumption was confirmed at an advisory board.[18]

Key: BAT, Best available therapy; ITC, indirect treatment comparison; JAKi, Janus Kinase inhibitor

B.3.6.2 Health-state unit costs and resource use

A summary of health state costs is provided in Table 82.

Table 82: List of health states and associated costs in the economic model

Health state	Sourced unit cost	Price year	Cost per event (adjusted to 2019)[125, 126]	Source
AML	£28,200 per year	2007	£32,087 per year	Wang et al., 2014,[139] Table 3
Palliative care	£760.38 per week	2015	**£813.72 per week**	Round et al., 2015[75] Table 5, sum of average health and social care costs for cancer patients
	£665.50 per week	2008	£804.83 per week	Addicott et al., 2008[76]
Key: AML, acute myeloid leukaemia. **Notes:** The costing option used in the base case model is indicated in bold				

Costs in the treatment health states are comprised of drug acquisition, drug administration, and resource use costs (described above), as well as adverse event costs (described below).

An all-encompassing cost for AML is assigned to patients while in the AML health state. The cost of AML was taken from a study by Wang et al. (2014) which considered medical costs of AML calculated using a micro-costing approach.[139] The micro-costing analysis included costs associated with treatment, hospitalisations, diagnostic tests, transfusions and associated complications.[139] A cost per life-month gained was generated, and this was converted to an annual cost for use in the model.

A cost for the palliative care state was identified from a study by Round et al. (2015) which estimated the average health and social care costs for cancer patients at the end of life.[75]

B.3.6.3 Summary of base-case analysis inputs

A summary of the variables and distributions applied in the economic model can be found in Appendix M.

B.3.6.4 Assumptions

Table 83 details the assumptions used in the economic model and their justification.

Table 83: Base case assumptions

Category	Assumption	Justification	Reference in submission
Response	'Spleen or symptom response' is the most appropriate definition for response	International Working Group Myeloproliferative Neoplasms Research and Treatment guidelines suggest both types of response should be considered.[51] This was also substantiated by clinical experts. [18]	B.3.3 - Types of response
	Patients can lose response but remain on treatment	Some clinical benefit may continue for the patient which does not meet the criteria for complete discontinuation. Therefore, the model accounts for this by estimating duration of response independently of treatment duration. This means that response is not artificially maintained for the entire treatment duration, which aims to reflect the clinical data.	B.3.3 - Duration of response
	Duration of response in the model was based on spleen response	Duration of response data for other response definitions were not collected for fedratinib or available for other treatments.	B.3.3 - Duration of response
Overall survival	The base case overall survival for fedratinib is based on the Gompertz extrapolation of the JAKARTA-2 Int-2/high risk survival data.	The Gompertz curve (Int-2/High-risk extrapolations) was selected as it closely aligned with clinicians' expectations for fedratinib survival	B.3.3 - Overall survival
	The most appropriate source for survival data and extrapolations is Schain et al. 2019.	From the available published data, Schain et al, 2019. Was indicated by clinicians as the population most representative of those expected to receive fedratinib in UK practice.	B.3.3 - Overall survival

Category	Assumption	Justification	Reference in submission
	The OS projected curves for fedratinib and BAT should not cross. Fedratinib OS is equal to BAT OS at the point of crossing.	Clinical opinion was that fedratinib OS would not be expected to be worse than BAT OS at any time point.	B.3.3 - Overall survival
Discontinuation	For patients receiving BAT (either as treatment arm or post-fedratinib), an explicit time-to-discontinuation is not estimated; it is assumed the patient remains on BAT until another event.	This is appropriate due to the lack of alternative treatment options.	B.3.3 - Discontinuation
BAT	Patients who receive fedratinib and discontinue to BAT do not receive ruxolitinib as part of BAT	Patients in this group would have received two JAKi treatments and therefore would not receive further ruxolitinib.	B.3.5 - Intervention and comparators' costs and resource use
	The composition of BAT is assumed to be equal to SIMPLIFY-2	SIMPLIFY-2 was identified as having the most realistic values for clinical practice by a UK advisory board.[18] This assumption is consistent across model inputs informing response and adverse events.	B.3.5 - Intervention and comparators' costs and resource use B.3.6.1 - Proportion of ruxolitinib within BAT
Utilities	MF-8D is an appropriate measure of utility for the selected patient population	MF-8D is a preference-based measure for MF which combines data from the MF-SAF and EORTC QLQ-C30 to generate utility scores. The MF-8D was the utility measure applied in the ruxolitinib NICE submission.[7]	B.3.4 - Mapping

Category	Assumption	Justification	Reference in submission
	Utility is dependent on patient response, as opposed to treatment arm. Additionally, response utility is derived using only spleen response as the predictor	A treatment-specific utility effect was not estimable in JAKARTA-2, given it was a single arm trial. Therefore, response was used to predict utility in line with NICE TA386. In contrast to NICE TA386, this submission allows patients receiving BAT to respond and therefore experience improvements in HRQoL- This is because SIMPLIFY-2 was used in the ITC, which showed a small improvement in response with BAT.	B.3.4 - Mapping
AML	It is appropriate to consider secondary AML as a health state	This is to reflect its prevalence and association with reduced life expectancy and health-related quality of life (HRQoL) in myelofibrosis.[74]	B.3.2 - Model structure
	Transition to AML is equal across treatment arms.	It is not clear whether treatment influences the rate of progression to AML.	Table 29: Implementation of events in the model
Key: AML, acute myeloid leukaemia; BAT, best available therapy; JAK, Janus Kinase; JAKi, JAK inhibitor; MF-8D, Myelofibrosis 8 dimensions; SE, standard error			

B.3.7. Base-case results

B.3.7.1 Base-case incremental cost-effectiveness analysis results

Table 84 presents the base case incremental cost-effectiveness results for fedratinib versus BAT at the net price. 1000 patients were used in the simulation, the convergence graphs for this analysis are in Appendix P.

The disaggregated clinical and economic outcomes by therapy arm and health state are presented in Appendix J.

B.3.8. Sensitivity analyses

B.3.8.1 Probabilistic sensitivity analysis

A probabilistic sensitivity analysis (PSA) was conducted in which all parameters were varied simultaneously over 1,000 iterations, by sampling their values from distributions (the convergence graphs for this analysis are in Appendix P). The results are summarised below in Table 85 and are also presented on a cost-effectiveness plane in Figure 28 and as a cost-effectiveness acceptability curve in Figure 29. All PSA iterations showed a positive QALY gain for fedratinib over BAT; 20.8% iterations reported that fedratinib resulted in a negative incremental cost for fedratinib. The probability of fedratinib being cost-effective is [redacted] at a willingness-to-pay (WTP) threshold of £30,000, and 97.9% at a WTP threshold of £50,000.

Table 84: Base-case results (based on net price)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	2.462	1.587	-	-	-	-	-
Fedratinib	[redacted]	3.309	2.202	8,545	0.848	0.615	13,905	13,905
Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.								

Table 85: Probabilistic sensitivity analysis results (based on net price)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	2.648	1.684	-	-	-	-	-
Fedratinib	[redacted]	3.546	2.308	6,480	0.898	0.624	10,384	10,384
Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.								

Figure 28: Cost-effectiveness plane – Fedratinib vs BAT

Figure 29: Cost-effectiveness acceptability curve – Fedratinib vs BAT

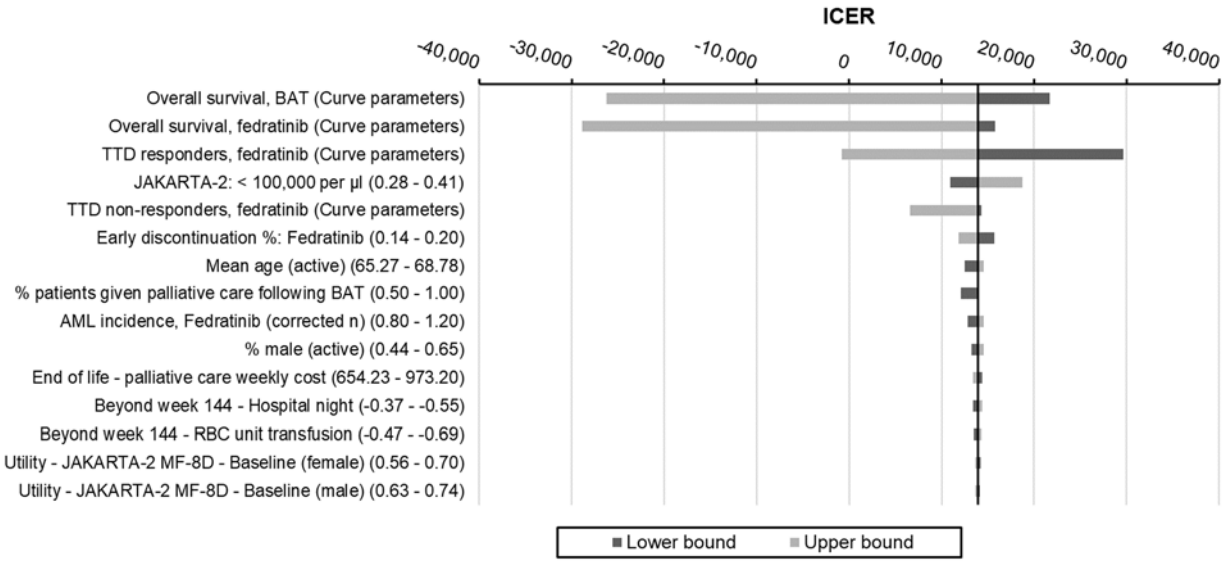
B.3.8.2 Deterministic sensitivity analysis

A series of one-way sensitivity analyses (OWSAs) were performed to evaluate the sensitivity of the model ICER to individual inputs, holding all else constant.

Confidence intervals, where available, were used to define the lower and upper bounds of a parameters. If a standard error (SE) was reported, bounds were set to ±1.96*SE. Alternatively, when uncertainty information were not available, lower and upper bounds were calculated based on the assumption that the SE was 10% of the mean deterministic value.

Figure 30 presents a tornado diagram which displays the 15 most influential parameters in descending order, in terms of their impact on the ICER at their lower/upper bounds. The parameters that most influenced the ICER were relating to patient OS for both fedratinib and BAT, this is expected given that OS is a key driver of LYs, and therefore QALYs and costs. The curve parameters for TTD for both responders and non-responders also significantly influence the ICER owing to the impact that patients remaining on therapy have on costs. The proportion of patients with a low platelet count also significantly impacts costs. These patients receive lower ruxolitinib dosing costs in BAT relative to patients which a high platelet count.

Figure 30: Results of one-way sensitivity analysis



B.3.8.3 Scenario analysis

Scenario analysis was performed to test some of the key inputs and assumptions to determine the effect upon the ICER. The top 15 scenarios that most effected the ICER are presented in Figure 31. The top five scenarios are further summarised in Table 86 with a description and rationale for each scenario. The scenarios that result in the largest impact on the ICER are those that test the OS modelling method used and BAT composition assumptions. These scenarios are included because there was no head-to-head data between fedratinib and BAT that would have informed the relative OS between the two arms and the composition of BAT. For the base case settings, clinical opinion was sought in determining the most appropriate approach for modelling OS and BAT composition.[18] With the BAT composition, the ICER is influenced by the proportion of ruxolitinib in BAT, this is because the absence of ruxolitinib significantly reduces costs without largely decreasing the QALY. However, the input of clinical opinion on the pathway for patients treated with ruxolitinib detailed in B.1.3.4 and B.3.5.1 justifies the use of the base case inputs, and how this assumption is used to produce overall model outcomes is described in Section B.3.6.1. The full scenario analysis results are presented in Appendix O.

Figure 31: Summary of modelling scenarios which had the most impact on the base case ICER

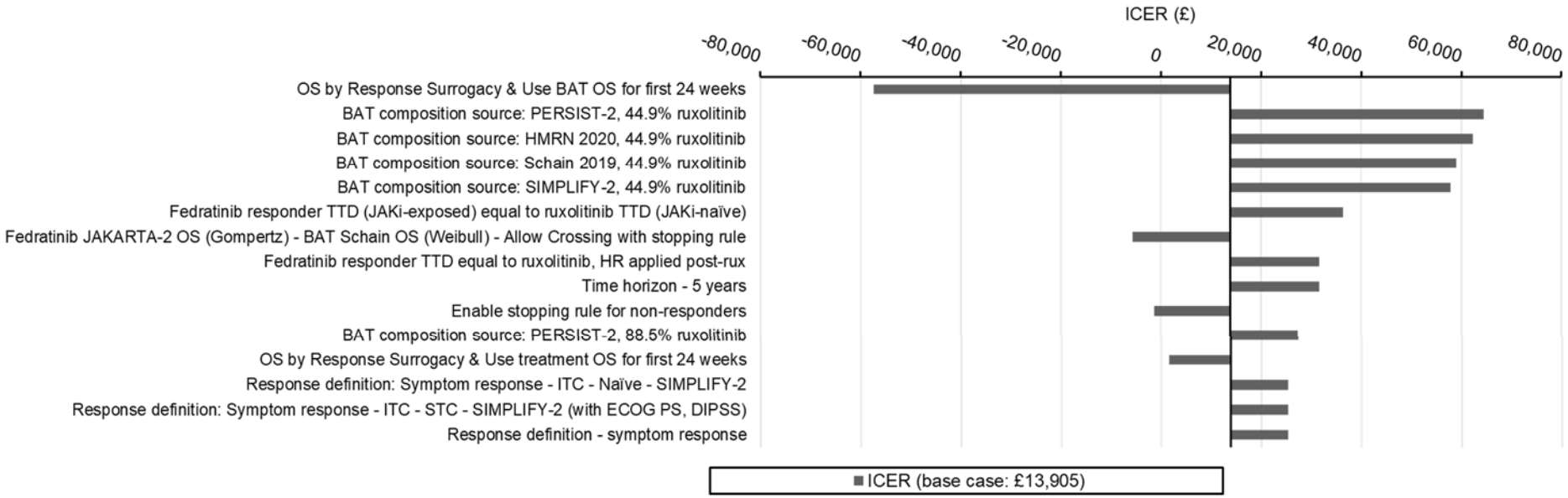


Table 86: Key scenario analysis (with net price)

Scenario and cross reference	Scenario detail	Brief rationale	Impact on base-case ICER
Base case			£13,905
OS by response surrogacy (Appendix L.7)	In this scenario, OS is assumed to be a factor of response. Non-responders experience non-responder OS and responders experience non-responder OS with a hazard ratio applied.[85]	OS is a primary driver of life-year and efficacy outcomes. Given the relative immaturity of the JAKARTA-2 fedratinib OS data, this scenario derived OS as a product of response based on surrogacy relationship reported in available literature.[83-85]	-£71,231 (Dominant)
BAT composition source: PERSIST-2, 44.9% ruxolitinib (B.3.5.1)	The BAT composition source is taken from PERSIST-2	The proportion of ruxolitinib is a key driver of drug costs in the model. Given the justification presented in Section B.3.6.1 it is reasonable to assume that there is a high proportion of patients receiving ruxolitinib after ruxolitinib failure. The proportion of ruxolitinib use in PERSIST-2 could act as a reasonable lower-bound value for the proportion of ruxolitinib use in BAT. These scenarios explore the proportion of BAT using the PERSIST-2 ruxolitinib proportion to test the assumptions of using different proportions of BAT in the model and the impact this has on incremental costs.	+£50,540
BAT composition source: HMRN 2020, 44.9% ruxolitinib (B.3.5.1)	The BAT composition source is taken from HMRN 2020 audit, with ruxolitinib % taken from PERSIST-2		+£48,403
BAT composition source: Schain 2019, 44.9% ruxolitinib (B.3.5.1)	The BAT composition source is taken from Schain et al 2019, with ruxolitinib % taken from PERSIST-2		+£45,064
BAT composition source: SIMPLIFY-2, 44.9% ruxolitinib (B.3.5.1)	The BAT composition source is taken from SIMPLIFY-2, with ruxolitinib % taken from PERSIST-2		+£43,918

Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; KM, Kaplan–Meier, OS, overall survival.

B.3.8.4 Summary of sensitivity analyses results

The probabilistic sensitivity analysis demonstrated that the conclusion that fedratinib is cost-effective versus BAT is highly likely. The CEAC, based upon 1000 PSA iterations, estimates that the probability of fedratinib being cost-effective at WTP threshold of £50,000 is 97.9%. The OWSA showed that the cost-effectiveness results were primarily sensitive to OS and TTD, which were derived from multivariate normal distributions as opposed to single variable parameters; it is not uncommon for curves varied this way to produce exaggerated or clinically implausible outcomes. Nevertheless, all parameters remained below a £30,000 WTP threshold.

A wide range of scenario analyses were performed on key model assumptions and alternative choices to test the robustness of the base case results. The scenarios showed that the assumptions surrounding the OS modelling assumptions and the composition of BAT had the most significant influence on the model outcomes. However, these assumptions have been validated during a clinical advisory board, indicating that these scenarios are likely to be clinically implausible or inappropriate for a UK setting. Of the 69 scenarios tested and 62 were below a WTP threshold of £30,000 and 65 were below a WTP threshold of £50,000.

B.3.9. Subgroup analysis

Subgroup analysis is not presented for this submission.

B.3.10. Validation

B.3.10.1 Validation of cost-effectiveness analysis

Expert clinical and health economic input was sought during the development of the cost-effectiveness model to ensure that the inputs and assumptions used in the base case analysis were relevant to UK clinical practice.

An advisory board was held in which the following model features were validated by both clinical and health economic experts, as appropriate:

- Model structure
- Clinical care pathway

- Relevant comparators
- Extrapolations for OS
- Composition of BAT

Once the model was finalised, technical validation was conducted by health economic modellers. A programmer (other than the one who built the model) reviewed all formulae, code and labelling in the model. Sensible lower and upper bounds (e.g. £0 for costs, but not negative costs) were input to the model one parameter at a time and the corresponding changes in the results were observed.

The results were checked against their expected impact. For example, setting all AE cost inputs to zero would result in AE cost outputs of £0 across both treatment arms.

B.3.11. Interpretation and conclusions of economic evidence

The economic analysis performed is based on a *de novo* economic model with a structure designed to reflect the disease indication of myelofibrosis in the most simplistic form while capturing the relevant outcomes. The model structure is based on the previous myelofibrosis technology appraisal of ruxolitinib (TA386), the first Janus kinase inhibitor (JAK-inhibitor) approved for myelofibrosis.

The model synthesises the most relevant and recently available efficacy and safety data from clinical trials and publications and used robust statistical techniques to establish the comparative efficacy and cost of fedratinib and BAT. Technology appraisal results and real-world evidence were used to inform resource use inputs, and clinical input was used to inform base case assumptions from clinical experts in myelofibrosis.

Results of the economic analyses indicate that fedratinib is a highly effective treatment for patients with MF who have been treated with ruxolitinib, even when comparing to a high proportion of patients who are continued on suboptimal ruxolitinib due to a lack of treatment options. Fedratinib provides 0.85 additional LYs and 0.61 additional QALYs versus BAT in those that have a median life expectancy of less than 2 years and no other treatment option.

Sensitivity analyses indicate that the cost-effectiveness results are robust and base case results suggest that fedratinib can be cost-effective at a threshold of £50,000 per QALY using the net price provided by Celgene, a BMS company.

This supports the argument that fedratinib is an innovative drug with a novel mechanism of action. Fedratinib selectively inhibits JAK2, with higher inhibitory activity for JAK2 over family members JAK1, JAK3 and TYK2 and is a more selective inhibitor of JAK2 than ruxolitinib. It is the only JAK inhibitor with demonstrable efficacy in a population that are relapsed, refractory, or intolerant to ruxolitinib who have a high unmet need.

The key limitations of the analysis are a lack of a head-to-head trial between fedratinib and BAT, which lead to the reliance of ITCs and naïve comparisons. Additionally, the clinical hold on fedratinib led to challenges in determining the true treatment effect, particularly for overall survival. Despite these challenges, fedratinib offers an alternative, convenient, well tolerated oral therapy that delivers clinically meaningful outcomes, including a survival gain for patients treated with ruxolitinib. Fedratinib, therefore, offers a step change in the clinical treatment pathway for myelofibrosis patients.

B.4. References

1. Mughal TI, Vaddi K, Sarlis NJ and Verstovsek S. Myelofibrosis-associated complications: pathogenesis, clinical manifestations, and effects on outcomes. *International journal of general medicine*. 2014; 7:89.
2. Song M-K, Park B-B and Uhm J-E. Understanding Splenomegaly in Myelofibrosis: Association with Molecular Pathogenesis. *International journal of molecular sciences*. 2018; 19(3):898.
3. Romano M, Sollazzo D, Trabanelli S, et al. Mutations in JAK2 and Calreticulin genes are associated with specific alterations of the immune system in myelofibrosis. *Oncoimmunology*. 2017; 6(10):e1345402.
4. Ihle JN and Gilliland DG. Jak2: normal function and role in hematopoietic disorders. *Current opinion in genetics & development*. 2007; 17(1):8-14.
5. Gangat N, Caramazza D, Vaidya R, et al. DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status. *Journal of Clinical Oncology*. 2011; 29(4):392-7.
6. Michaelis LC. Risk stratification in Myelofibrosis: The quest for simplication. *Haematologica*. 2017; 102(1):2-3.
7. National Institute of Care Excellence (NICE). Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. Technology appraisal guidance [TA386]. 2016. (Updated: 23 March 2016) Available at: https://www.nice.org.uk/guidance/TA386/chapter/1-Recommendations. Accessed: 13 November 2019.
8. DeAngelo D. Diagnosis/risk classification for primary myelofibrosis. 2017. Available at: https://www.targetedonc.com/case-based-peer-perspectives/myeloproliferative-neoplasms/deangelo-highrisk-mf/diagnosis-risk-classification-for-primary-myelofibrosis. Accessed: 5 December 2019.
9. Tefferi A, Lasho TL, Jimma T, et al. One thousand patients with primary myelofibrosis: the mayo clinic experience. *Mayo Clinic Proceedings*. Elsevier, 2012, p. 25-33.
10. Quintás-Cardama A, Kantarjian H, Pierce S, et al. Prognostic model to identify patients with myelofibrosis at the highest risk of transformation to acute myeloid leukemia. *Clinical Lymphoma Myeloma and Leukemia*. 2013; 13(3):315-8. e2.
11. Mitra D, Kaye JA, Piecoro LT, et al. Symptom burden and splenomegaly in patients with myelofibrosis in the United States: a retrospective medical record review. *Cancer medicine*. 2013; 2(6):889-98.
12. Haematological Malignancy Research Network (HMRN). Haematological Malignancy Research Network: Clinical Management, Resource Utilisation and Outcome in Primary and Secondary Myelofibrosis. 27 May 2020 2020.
13. Cervantes F, Dupriez B, Pereira A, et al. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood*. 2009; 113(13):2895-901.
14. Passamonti F, Cervantes F, Vannucchi AM, et al. A dynamic prognostic model to predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment). *Blood*. 2010; 115(9):1703-8.

15. Schain F, Vago E, Song C, et al. Survival outcomes in myelofibrosis patients treated with ruxolitinib: A population-based cohort study in Sweden and Norway. *Eur J Haematol*. 2019; 103(6):614-9.
16. Kuykendall AT, Shah S, Talati C, et al. Between a rux and a hard place: evaluating salvage treatment and outcomes in myelofibrosis after ruxolitinib discontinuation. *Ann Hematol*. 2018; 97(3):435-41.
17. Palandri F, Breccia M, Bonifacio M, et al. Life after ruxolitinib: Reasons for discontinuation, impact of disease phase, and outcomes in 218 patients with myelofibrosis. *Cancer*. 2019.
18. Celgene. Myelofibrosis Advisory Board Meeting Report. April 8th 2020.
19. Scherber R, Dueck AC, Johansson P, et al. The Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF): international prospective validation and reliability trial in 402 patients. *Blood*. 2011; 118(2):401-8.
20. Mesa R, Miller CB, Thyne M, et al. Myeloproliferative neoplasms (MPNs) have a significant impact on patients' overall health and productivity: the MPN Landmark survey. *BMC cancer*. 2016; 16(1):167.
21. Iurlo A, Cattaneo D and Gianelli U. Blast Transformation in Myeloproliferative Neoplasms: Risk Factors, Biological Findings, and Targeted Therapeutic Options. *International journal of molecular sciences*. 2019; 20(8):1839.
22. Polverelli N, Breccia M, Benevolo G, et al. Risk factors for infections in myelofibrosis: role of disease status and treatment. A study on 507 patients. American Society of Hematology Washington, DC, 2015.
23. Devendra K, Falchi L and Verstovsek S. The underappreciated risk of thrombosis and bleeding in patients with myelofibrosis: a review. *Annals of hematology*. 2017; 96(10):1595-604.
24. Kander EM, Raza S, Zhou Z, et al. Bleeding complications in BCR-ABL negative myeloproliferative neoplasms: prevalence, type, and risk factors in a single-center cohort. *International Journal of Hematology*. 2015; 102(5):587-93.
25. Finazzi G, Carobbio A, Thiele J, et al. Incidence and risk factors for bleeding in 1104 patients with essential thrombocythemia or prefibrotic myelofibrosis diagnosed according to the 2008 WHO criteria. *Leukemia*. 2012; 26(4):716-9.
26. Mesa RA, Gotlib J, Gupta V, et al. Effect of ruxolitinib therapy on myelofibrosis-related symptoms and other patient-reported outcomes in COMFORT-I: a randomized, double-blind, placebo-controlled trial. *Journal of Clinical Oncology*. 2013; 31(10):1285.
27. Harrison CN, Koschmieder S, Foltz L, et al. The impact of myeloproliferative neoplasms (MPNs) on patient quality of life and productivity: results from the international MPN Landmark survey. *Annals of hematology*. 2017; 96(10):1653-65.
28. Yu J, Parasuraman S, Paranagama D, et al. Impact of Myeloproliferative neoplasms on patients' employment status and work productivity in the United States: results from the living with MPNs survey. *BMC cancer*. 2018; 18(1):420.
29. Harrison CN, Mesa RA, Kiladjian JJ, et al. Health-related quality of life and symptoms in patients with myelofibrosis treated with ruxolitinib versus best available therapy. *Br J Haematol*. 2013; 162(2):229-39.
30. Nolte S, Liegl G, Petersen MA, et al. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. *Eur J Cancer*. 2019; 107:153-63.

31. Daghia G, Zabelina T, Zeck G, et al. Allogeneic stem cell transplantation for myelofibrosis patients aged >/=65 years. *Eur J Haematol*. 2019; 103(4):370-8.
32. Vannucchi AM and Harrison CN. Emerging treatments for classical myeloproliferative neoplasms. *Blood, The Journal of the American Society of Hematology*. 2017; 129(6):693-703.
33. Kroger N and Mesa RA. Choosing between stem cell therapy and drugs in myelofibrosis. *Leukemia*. 2008; 22(3):474-86.
34. Novartis. Jakavi (ruxolitinib) Summary of product characteristics. 2019. Available at: https://www.ema.europa.eu/en/documents/product-information/jakavi-epar-product-information_en.pdf. Accessed: 11 March 2020.
35. Vannucchi AM, Barbui T, Cervantes F, et al. Philadelphia chromosome-negative chronic myeloproliferative neoplasms: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2015; 26 Suppl 5:v85-99.
36. Harrison C, Kiladjian JJ, Al-Ali HK, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N Engl J Med*. 2012; 366(9):787-98.
37. Harrison CN, Schaap N, Vannucchi AM, et al. Janus kinase-2 inhibitor fedratinib in patients with myelofibrosis previously treated with ruxolitinib (JAKARTA-2): a single-arm, open-label, non-randomised, phase 2, multicentre study. *Lancet Haematol*. 2017; 4(7):e317-e24.
38. Pardanani A and Tefferi A. Definition and management of ruxolitinib treatment failure in myelofibrosis. *Blood Cancer J*. 2014; 4(12):e268.
39. Ragheb M, Harrison CN and McLornan DP. Current and future role of fedratinib in the treatment of myelofibrosis. *Future Oncology*. 2020; 16(6):175-86.
40. Tefferi A and Pardanani A. Serious adverse events during ruxolitinib treatment discontinuation in patients with myelofibrosis. *Mayo Clinic Proceedings*. Elsevier, 2011, p. 1188-91.
41. Mascarenhas J, Hoffman R, Talpaz M, et al. Pacritinib vs Best Available Therapy, Including Ruxolitinib, in Patients With Myelofibrosis: A Randomized Clinical Trial. *JAMA Oncol*. 2018; 4(5):652-9.
42. Harrison CN, Vannucchi AM, Platzbecker U, et al. Momelotinib versus best available therapy in patients with myelofibrosis previously treated with ruxolitinib (SIMPLIFY 2): a randomised, open-label, phase 3 trial. *Lancet Haematol*. 2018; 5(2):e73-e81.
43. Celgene. ARD12181 (JAKARTA2): A Phase II, multicenter, open-label, single-arm study of SAR302503 in patients previously treated with ruxolitinib and with a current diagnosis of intermediate or high-risk primary myelofibrosis, post-polycythemia vera myelofibrosis, or post-essential thrombocythemia myelofibrosis (clinical study report). 18 Dec 2018 2018.
44. Harrison C, Schaap N, Vannucchi A, et al. Fedratinib in patients with myeloproliferative neoplasm-associated myelofibrosis previously treated with ruxolitinib: a reanalysis of the Phase 2 JAKARTA-2 study. European Hematology Association. Amsterdam, Netherlands. June 13–16 2019. PS1459.
45. Harrison CN, Mesa RA, Jamieson C, et al. Case series of potential Wernicke's encephalopathy in patients treated with Fedratinib. *Blood*. 2017; 130(Supplement 1):4197-.
46. Harrison CN, Schaap N, Vannucchi AM, et al. Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the JAKARTA2 study using stringent criteria for ruxolitinib failure. *American Journal of Hematology*. 2020.

47. Celgene. Global Fedratinib Strategy Advisory Board. April 18 2018.
48. Pardanani A, Harrison C, Cortes JE, et al. Safety and efficacy of fedratinib in patients with primary or secondary myelofibrosis: a randomized clinical trial. *JAMA Oncol*. 2015; 1(5):643-51.
49. Celgene. Overall survival estimates from JAKARTA and JAKARTA-2. 2019.
50. Celgene. JAKARTA-2 patient level data analysis 2020.
51. Tefferi A, Cervantes F, Mesa R, et al. Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. *Blood*. 2013; 122(8):1395-8.
52. Verstovsek S, Kantarjian H, Mesa RA, et al. Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *New England Journal of Medicine*. 2010; 363(12):1117-27.
53. Verstovsek S, Mesa RA, Gotlib J, et al. Three-year efficacy, overall survival, and safety of ruxolitinib therapy in patients with myelofibrosis from the COMFORT-I study. *Haematologica*. 2015:haematol. 2014.115840.
54. Miller CB, Komrokji RS, Mesa RA, et al. Practical measures of clinical benefit with ruxolitinib therapy: an exploratory analysis of COMFORT-I. *Clinical Lymphoma Myeloma and Leukemia*. 2017; 17(8):479-87.
55. Khan I, Bashir Z and Forster M. Interpreting small treatment differences from quality of life data in cancer trials: an alternative measure of treatment benefit and effect size for the EORTC-QLQ-C30. *Health Qual Life Outcomes*. 2015; 13:180.
56. Celgene. Indirect treatment comparisons of fedratinib versus best available therapy for Janus Kinase-inhibitor treated patients with myelofibrosis. 18 September 2019 2019.
57. Phillippo D, Ades T, Dias S, et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. 2016. Available at: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/05/Population-adjustment-TSD-FINAL.pdf. Accessed: To complete.
58. Albert A and Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 1984; 71(1):1-10.
59. Allison PD. Convergence failures in logistic regression. *SAS Global Forum*. 2008, p. 1-11.
60. Mesa RA, Kiladjian J-J, Verstovsek S, et al. Comparison of placebo and best available therapy for the treatment of myelofibrosis in the phase 3 COMFORT studies. *Haematologica*. 2014; 99(2):292-8.
61. Rojas R, Balmaceda C, Vargas C and Espinoza MA. Cost effectiveness of ruxolitinib versus best available therapy for the treatment of myelofibrosis from public health perspective in Chile. *Value Health*. 2016; 19(7):A590.
62. Vandewalle B, Andreozzi V, Almeida J and Félix J. Pharmacoeconomics of ruxolitinib therapy in patients with myelofibrosis. *Journal of Medical Economics*. 2016; 19(4):424-31.
63. Hahl J, Kurki S, Miettinen T and Snicker K. Cost-effectiveness of ruxolitinib for the treatment of myelofibrosis in Finland. Economic evaluation based on finnish auria biobank data on health care resource utilization. *Value Health*. 2015; 18(7):A669.
64. PBAC[Ruxolitinib]. Ruxolitinib; 5 mg tablet 56, 15 mg tablet 56, 20 mg tablet 56; Jakavi®. 2015. (Updated: 03 July 2015) Available at:

http://www.pbs.gov.au/info/industry/listing/elements/pbac-meetings/psd/2015-03/ruxolitinib-jakavi-psd-03-2015. Accessed: 14 November 2018.
65. SMC[Ruxolitinib]. Ruxolitinib (as phosphate), 5mg, 15mg, & 20mg tablets (Jakavi®) SMC No. (867/13). 2015. (Updated: 09 March 2015) Available at: https://www.scottishmedicines.org.uk/medicines-advice/ruxolitinib-jakavi-fullsubmission-86713/. Accessed: 02 November 2018.
66. CADTH[Ruxolitinib]. Jakavi® for myelofibrosis. 2013. (Updated: 14 January 2013) Available at: https://www.cadth.ca/jakavi-myelofibrosis. Accessed: 10 October 2018.
67. NCPE[Ruxolitinib]. Cost-effectiveness of ruxolitinib (Jakavi®) for the treatment of splenomegaly or disease-related symptoms in adult patients with primary myelofibrosis, post olycythaemia vera melofibrosis or post essential hrombocythaemia myelofibrosis. 2013. Available at: http://www.ncpe.ie/drugs/ruxolitinib-jakavi/. Accessed: 02 November 2018.
68. Khalid El Ouagari, Christopher John Knight and Mendelson ET. Cost-Effectiveness of Ruxolitinib Versus Best-Available Therapy for Medical Treatment of Myelofibrosis: Canadian Societal Perspective. Blood. 16 November 2012 2012. 901.
69. Clinicaltrials.gov. Controlled Myelofibrosis Study With Oral Janus-associated Kinase (JAK) Inhibitor Treatment-II: The COMFORT-II Trial. 2019. (Updated: 19 August 2019) Available at: https://clinicaltrials.gov/ct2/show/NCT00934544. Accessed: 27 March 2020.
70. Clinicaltrials.gov. COntrolled MyeloFibrosis Study With ORal JAK Inhibitor Treatment: The COMFORT-I Trial. 2018. (Updated: 12 March 2018) Available at: https://clinicaltrials.gov/ct2/show/NCT00952289. Accessed: 27 March 2020.
71. Clinicaltrials.gov. Open Label Ruxolitinib (INCB018424) in Patients With Myelofibrosis and Post Polycythemia Vera/Essential Thrombocythemia Myelofibrosis. 2018. Available at: https://clinicaltrials.gov/ct2/show/NCT00509899. Accessed: 27 March 2020.
72. Celgene. ARD12181 (JAKARTA2): A Phase II, Multicenter, Open-Label, Single-Arm Study of SAR302503 in Subjects with a Current Diagnosis of Intermediate or High-Risk Primary Myelofibrosis, Post Polycythemia Vera Myelofibrosis, or Post Essential Thrombocythemia Myelofibrosis - CSR Addendum. (Clinical Study Report) 14 December 2018 2018.
73. European LeukemiaNet (ELN). Developing stratified diagnostic and treatment approaches. 2012. Available at: https://www.leukemia-net.org/content/physicians/recommendations/e9824/infoboxContent9825/EuropeanLeukemiaNet-Developingstratifieddiagnosticandtreatmentapproaches_Update2013_ohnePublikationen.pdf. Accessed: 13 March 2020.
74. Mesa RA, Li CY, Ketterling RP, et al. Leukemic transformation in myelofibrosis with myeloid metaplasia: a single-institution experience with 91 cases. *Blood*. 2005; 105(3):973-7.
75. Round J, Jones L and Morris S. Estimating the cost of caring for people with cancer at the end of life: A modelling study. *Palliat Med*. 2015; 29(10):899-907.
76. Addicott R and Dewar S. Improving choice at end of life: a descriptive analysis of the impact and costs of the Marie Curie delivering choice programme in Lincolnshire. 2008. Available at: https://www.kingsfund.org.uk/sites/default/files/improving-choice-end-of-life-

descriptive-analysis-impact-costs-marie-curie-choice-programme-lincolnshire-rachael-addicot-steve-dewar-april-2008.pdf. Accessed: 6 March 2020.
77. Davis S, Stevenson M, Tappenden P and Wailoo AJ. NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation. 2014. 2014. Available at: http://www.nicedsu.org.uk. Accessed: 13 March 2020.
78. Mukuria C, Rowen D, Brazier JE, et al. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015; 18(6):846-55.
79. Reilly JT, McMullin MF, Beer PA, et al. Guideline for the diagnosis and management of myelofibrosis. *Br J Haematol*. 2012; 158(4):453-71.
80. McLornan DP and Harrison CN. Guidance on changing therapy choice in myelofibrosis. *Blood Adv*. 2020; 4(4):607-10.
81. Harrison CN, Schaap N and Mesa RA. Management of myelofibrosis after ruxolitinib failure. *Ann Hematol*. 2020; 99(6):1177-91.
82. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. . 2011. (Updated: 2013) Available at: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf.
83. Verstovsek S, Kantarjian HM, Estrov Z, et al. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. *Blood*. 2012; 120(6):1202-9.
84. Vannucchi AM, Kantarjian HM, Kiladjian JJ, et al. A pooled analysis of overall survival in COMFORT-I and COMFORT-II, 2 randomized phase III trials of ruxolitinib for the treatment of myelofibrosis. *Haematologica*. 2015; 100(9):1139-45.
85. Palandri F, Palumbo GA, Bonifacio M, et al. Baseline factors associated with response to ruxolitinib: an independent study on 408 patients with myelofibrosis. *Oncotarget*. 2017; 8(45):79073-86.
86. Celgene [data on file]. External clinical validation exercises - fedratinib. 2019.
87. Mascarenhas J, Komrokji RS, Cavo M, et al. Imetelstat Is Effective Treatment for Patients with Intermediate-2 or High-Risk Myelofibrosis Who Have Relapsed on or Are Refractory to Janus Kinase Inhibitor Therapy: Results of a Phase 2 Randomized Study of Two Dose Levels. 2018. Available at: https://www.geron.com/file.cfm/53/docs/imetelstat_Imbark_ASH%202018_FINAL.pdf . Accessed: 10 October 2019.
88. Gupta V, Kosiorek HE, Klisovic RB, et al. Exploring the Potential of JAK1/2 Inhibitor Ruxolitinib with Reduced Intensity Hematopoietic Cell Transplantation (HCT) for Myelofibrosis: Stage I Results of a Prospective Trial Conducted through the Myeloproliferative Disorders - Research Consortium (MPD-RC). *Blood*. 2016; 128(22):1126-.
89. Kadir SSSA, Christopeit M, Wulf G, et al. Impact of ruxolitinib pretreatment on outcomes after allogeneic stem cell transplantation in patients with myelofibrosis. *European Journal of Haematology*. 2018; 101(3):305-17.
90. Mehra M, Potluri R, He J, et al. Characterization of disease, treatment patterns, and outcomes of patients with myelofibrosis: analysis of two United States commercial claims databases. *American Society of Hematology (ASH) 58th Annual Meeting*. San Diego, California2016.

91. Miller K, Hobbs G, Ho VT, et al. The effect of pre-transplant JAK 1/2 inhibitors on outcomes of myelofibrosis patients who receive allogeneic stem cell transplant. *Journal of Clinical Oncology*. 2018; 36(15_suppl):7072-.
92. Newberry KJ, Patel K, Masarova L, et al. Clonal evolution and outcomes in myelofibrosis after ruxolitinib discontinuation. *Blood*. 2017; 130(9):1125-31.
93. Palandri F, Elli E, Polverelli N, et al. Outcome of Patients with Myelofibrosis after Ruxolitinib Failure: Role of Disease Status and Treatment Strategies in 214 Patients. *Blood*. 2018; 132(Suppl 1):4277.
94. Shanavas M, Popat U, Michaelis LC, et al. Outcomes of Allogeneic Hematopoietic Cell Transplantation in Patients with Myelofibrosis with Prior Exposure to Janus Kinase 1/2 Inhibitors. *Biol Blood Marrow Transplant*. 2016; 22(3):432-40.
95. Rowen D, Brazier J, Young T, et al. Deriving a Preference-Based Measure for Cancer Using the EORTC QLQ-C30. *Value in Health*. 2011; 14(5):721-31.
96. Rafia R and Mukuria CW. Exploratory Psychometric Analysis of The Eq-5d In A Myelofibrosis Population. *Value in Health*. 2015; 18(3):A211.
97. Mukuria CW, Brazier J and Rafia R. Does the generic cancer outcome measure eortc qlq-c30 work in myelofibrosis? *Value in Health*. 2015; 18(3):A210-A1.
98. Tielemans MM, Jaspers Focks J, van Rossum LG, et al. Gastrointestinal symptoms are still prevalent and negatively impact health-related quality of life: a large cross-sectional population based study in The Netherlands. *PLoS One*. 2013; 8(7):e69876.
99. Hollingworth W, Gray DT, Martin BI, et al. Rapid magnetic resonance imaging for diagnosing cancer-related low back pain. *J Gen Intern Med*. 2003; 18(4):303-12.
100. Beusterien KM, Davies J, Leach M, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. *Health Qual Life Outcomes*. 2010; 8:50.
101. Doyle S, Lloyd A and Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer*. 2008; 62(3):374-80.
102. Schremser K, Rogowski WH, Adler-Reichel S, et al. Cost-Effectiveness of an Individualized First-Line Treatment Strategy Offering Erlotinib Based on EGFR Mutation Testing in Advanced Lung Adenocarcinoma Patients in Germany. *Pharmacoeconomics*. 2015; 33(11):1215-28.
103. Lachaine J, Mathurin K, Barakat S and Couban S. Economic evaluation of arsenic trioxide compared to all-trans retinoic acid + conventional chemotherapy for treatment of relapsed acute promyelocytic leukemia in Canada. *Eur J Haematol*. 2015; 95(3):218-29.
104. Nafees B, Stafford M, Gavriel S, et al. Health state utilities for non small cell lung cancer. *Health Qual Life Outcomes*. 2008; 6:84.
105. Pan F, Peng S, Fleurence R, et al. Economic Analysis of Decitabine Versus Best Supportive Care in the Treatment of Intermediate- and High-Risk Myelodysplastic Syndromes From a US Payer Perspective. *Clinical Therapeutics*. 2010; 32(14):2444-56.
106. Forsythe A, Brandt PS, Dolph M, et al. Systematic review of health state utility values for acute myeloid leukemia. *ClinicoEconomics and outcomes research : CEOR*. 2018; 10:83-92.
107. Färkkilä N, Torvinen S, Roine RP, et al. Health-related quality of life among breast, prostate, and colorectal cancer patients with end-stage disease. *Quality of*

life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2014; 23(4):1387-94.
108. Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value in Health*. 2016; 19(6):704-19.
109. Ara R and Brazier JE. Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. *Value in Health*. 2010; 13(5):509-18.
110. Monthly Index of Medical Specialities (MIMS). Drugs. 2020. Available at: https://www.mims.co.uk/drugs. Accessed: 4 March 2020.
111. Gov.UK. Drugs and pharmaceutical electronic market information tool (eMIT). 2019. (Updated: 14 November 2019) Available at: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit. Accessed: 4 March 2020.
112. Monthly Index of Medical Specialities (MIMS). Jakavi. 2020. Available at: https://www.mims.co.uk/drugs/cancer/antineoplastics/jakavi. Accessed: 26 March 2020.
113. National Health Servive Business Services Authority (NHSBSA). Drug Tariff. 2020. (Updated: March 2020) Available at: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff. Accessed: 4 March 2020.
114. Monthly Index of Medical Specialities (MIMS). Danazol. 2020. Available at: https://www.mims.co.uk/drugs/obstetrics-and-gynaecology/breast-disorders/danazol. Accessed: 4 March 2020.
115. Monthly Index of Medical Specialities (MIMS). Thalidomide Celgene. 2020. Available at: https://www.mims.co.uk/drugs/cancer/antineoplastics/thalidomide-celgene. Accessed: 4 March 2020.
116. Monthly Index of Medical Specialities (MIMS). Cytarabine. 2020. Available at: https://www.mims.co.uk/drugs/cancer/antineoplastics/cytarabine. Accessed: 4 March 2020.
117. Monthly Index of Medical Specialities (MIMS). Dacogen. 2020. Available at: https://www.mims.co.uk/drugs/cancer/antineoplastics/dacogen. Accessed: 4 March 2020.
118. Monthly Index of Medical Specialities (MIMS). Roferon-A. 2020. Available at: https://www.mims.co.uk/drugs/infections-and-infestations/viral-infections/roferon-a. Accessed: 6 March 2020.
119. Monthly Index of Medical Specialities (MIMS). Pegasys. 2020. Available at: https://www.mims.co.uk/drugs/infections-and-infestations/viral-infections/pegasys. Accessed: 4 March 2020.
120. National Health Service (NHS). National schedule of NHS costs 2018/19. 2020. (Updated: 19 February 2020) Available at: https://improvement.nhs.uk/resources/national-cost-collection/. Accessed: 4 March 2020.
121. Haematological Malignancy Research Network (HMRN). Haematological Malignancy Research Network: Myelofibrosis Audit. 2012. (Updated: 2 October 2012) Available at: https://www.hmrn.org/Download.aspx?target=document&id=3. Accessed: 6 March 2020.
122. Mead AJ, Milojkovic D, Knapper S, et al. Response to ruxolitinib in patients with intermediate-1-, intermediate-2-, and high-risk myelofibrosis: results of the UK ROBUST Trial. *Br J Haematol*. 2015; 170(1):29-39.

123. Al-Ali HK, Griesshammer M, le Coutre P, et al. Safety and efficacy of ruxolitinib in an open-label, multicenter, single-arm phase 3b expanded-access study in patients with myelofibrosis: a snapshot of 1144 patients in the JUMP trial. *Haematologica*. 2016; 101(9):1065-73.
124. Martino B, Coutre Pl, Griesshammer M, et al. Safety and Efficacy of Ruxolitinib in an Open-Label, Multicenter, Single-Arm, Expanded-Access Study in Patients with Myelofibrosis (MF): An 1144-Patient Update. *Blood*. 2014; 124(21):3197.
125. Curtis LA and Burns A. *Unit Costs of Health and Social Care 2019*. Kent, UK2019, p. 179.
126. Curtis LA. Unit Costs of Health and Social Care 2011. 2011. Available at: https://www.pssru.ac.uk/pub/uc/uc2011/uc2011.pdf. Accessed: 6 March 2020.
127. National Health Service (NHS). Private Patient Service at Dorset County Hospital. 2019. (Updated: 1 April 2019) Available at: https://www.dchft.nhs.uk/patients/private-patient-service/Documents/Private%20Patient%20Tariff%202019-2020.pdf. Accessed: 6 March 2020.
128. Varney S and Guest J. The Annual Cost to Society of Blood Transfusions in the UK. 2003. Available at: https://www.catalyst-health.co.uk/wp-content/uploads/2010/12/annual-cost-of-blood-transfusions-cropped.pdf. Accessed: 6 March 2020.
129. Celgene. Summary of Product Characteristics (SPC). Inrebic® (fedratinib) 100mg hard capsules. 2020. Available at: TBD. Accessed: TBD.
130. National Health Service (NHS). Provider to provider services. 2019. Available at: https://www.uclh.nhs.uk/aboutus/wwd/Documents/Provider%20to%20Provider%20Tariff%202018-19.pdf. Accessed: 27 May 2020.
131. Monthly Index of Medical Specialities (MIMS). Thiamine. 2020. Available at: https://www.mims.co.uk/drugs/nutrition/vitamin-supplements/thiamine. Accessed: 27 May 2020.
132. National Institute of Care Excellence (NICE). Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel-containing regimen [TA316]. 2014. Available at: https://www.nice.org.uk/guidance/ta316. Accessed: 6 March 2020.
133. Monthly Index of Medical Specialities (MIMS). Clarithromycin. 2020. Available at: https://www.mims.co.uk/drugs/infections-and-infestations/bacterial-infections/clarithromycin. Accessed: 6 March 2020.
134. Monthly Index of Medical Specialities (MIMS). Loperamide. 2020. Available at: https://www.mims.co.uk/drugs/gastrointestinal-tract/diarrhoea/loperamide. Accessed: 6 March 2020.
135. McCrone P, Seed PT, Dowson AJ, et al. Service use and costs for people with headache: a UK primary care study. *J Headache Pain*. 2011; 12(6):617-23.
136. Monthly Index of Medical Specialities (MIMS). Ondansetron. 2020. Available at: https://www.mims.co.uk/drugs/central-nervous-system/nausea-vomiting/ondansetron. Accessed: 25 June 2020.
137. Woods B, Hawkins N, Dunlop W, et al. Bendamustine versus chlorambucil for the first-line treatment of chronic lymphocytic leukemia in England and Wales: a cost-utility analysis. *Value Health*. 2012; 15(5):759-70.

138. Oncology S. Targeted Hematology and Oncology Therapeutics. 2018. Available at: http://investor.sierraoncology.com/download/Sierra+Analyst+%26+Investor+Call+Presentation_Final.pdf Accessed: 22 June 2020.
139. Wang HI, Aas E, Howell D, et al. Long-term medical costs and life expectancy of acute myeloid leukemia: a probabilistic decision model. *Value Health*. 2014; 17(2):205-14.

B.5. Appendices

Appendix C: Summary of product characteristics (SmPC) and European public assessment report (EPAR)

Appendix D: Identification, selection and synthesis of clinical evidence

Appendix E: Subgroup analysis

Appendix F: Adverse reactions

Appendix G: Published cost-effectiveness studies

Appendix H: Health-related quality-of-life studies

Appendix I: Cost and healthcare resource identification, measurement and valuation

Appendix J: Clinical outcomes and disaggregated results from the model

Appendix K: Checklist of confidential information

Appendix L: Supportive materials for the economic evaluation

Appendix M: Summary of base case analysis inputs

Appendix N: Supportive data from the HMRN analysis

Appendix O: Full sensitivity analysis results

Appendix P: Cost-effectiveness model convergence graphs

Addendum to NICE submission

Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis [ID1501]

Addendum overview

As part of the European Medicines Agency (EMA) regulatory submission for fedratinib, the Committee for Medicinal Products for Human Use (CHMP) requested that patients should be counted as non-responders if they achieved a ≥ 35% reduction in spleen volume from baseline to the End of Cycle 6 (spleen response – primary endpoint) with a dose higher than 400 mg/daily of fedratinib.

Further to the ID1501 stakeholder response submitted to NICE on 5 November 2020 by Celgene, a BMS company, this addendum has been prepared to describe analyses and model updates that were performed to align with the CHMP request. In addition, the addendum aims to provide further clarity on key matters of uncertainty following the technical engagement stage.

Finally, in acknowledgement of the uncertainty in the existing evidence base, the proposed net price of fedratinib has been updated to £[redacted] per pack. This change to the simple patient access scheme has been sent to PASLU.

1. Why fedratinib is a suitable candidate for the Cancer Drugs Fund

Unmet need in the previously treated with ruxolitinib setting

In the current pathway of clinical care, ruxolitinib is the only treatment that is specifically licensed or reimbursed by NICE for the treatment of myelofibrosis. As outlined within the original submission document, the lack of treatment options available means that there is a high unmet need for UK patients with myelofibrosis who have been previously treated with ruxolitinib. This is also recognised by the Evidence Review Group (ERG)[1, 2].

Outcomes in patients no longer responding to ruxolitinib are poor, with a loss of response associated with worsened symptoms and an increased spleen size, resulting in detriments to health-related quality of life (HRQoL). The survival outcomes in patients who have been treated with ruxolitinib are poor, with studies indicating a median OS of 13–16 months[5-8] following ruxolitinib discontinuation, though this may be as short as 6 months in some patients.[9]

Despite the poor prognosis and high symptom burden in these patients, the standard of care is currently limited to best available therapy (BAT), which is not associated with significant reductions in spleen volume or total symptom scores.[10] In patients that do respond to ruxolitinib, many will become relapsed or refractory to ruxolitinib over time. With no alternative effective treatment options which can significantly reduce the spleen volume or total symptom score, many relapsed and refractory patients remain on suboptimal ruxolitinib therapy.[3, 4] These patients who continue suboptimal ruxolitinib have poor outcomes, including survival as stated by treating clinicians. [3]

This demonstrates a clear unmet need for a new, efficacious, and tolerable treatment option for patients who are relapsed, refractory or intolerant to ruxolitinib, so that these poor outcomes can be improved. Fedratinib, which was granted a marketing authorisation by the EMA on the 8 February 2021 for the treatment of disease-related splenomegaly or symptoms in adult patients with primary myelofibrosis, post polycythaemia vera myelofibrosis or post essential thrombocythaemia myelofibrosis who are Janus Associated Kinase (JAK) inhibitor naïve or have been treated with ruxolitinib, can address this clear unmet need.

Fedratinib, a targeted and novel JAK-2 inhibitory therapy, offers an effective treatment that has shown clinically meaningful spleen and symptom responses in patients who have been treated with ruxolitinib.[11] These benefits can lead to considerable HRQoL and survival improvements, in a patient population that would otherwise experience poor outcomes.

The proposed position of fedratinib in the UK treatment pathway is narrower than the marketing authorisation because the population of patients who have been treated with ruxolitinib represents the greatest unmet need in myelofibrosis, and for which the clinical and cost-effectiveness of fedratinib is most demonstrable.

Uncertainty

Celgene believes fedratinib would make a good candidate to briefly enter the Cancer Drugs Fund (CDF) with the majority of the uncertainties being resolved with the ongoing FREEDOM-2 trial and/or the CDF.

FREEDOM-2 is a phase III, randomised, controlled trial comparing fedratinib with BAT in patients with DIPSS intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib. The trial will provide data on the following:

1. Relative efficacy of fedratinib compared to BAT, which will include ruxolitinib.
2. Quality of life data.
3. Survival data of those randomised to fedratinib and BAT. Although due to cross-over, some uncertainty may remain.
4. Dosing of ruxolitinib in the setting of BAT, therefore providing greater clarity on drug cost.
5. Discontinuation of fedratinib.
6. Proportions of therapy within BAT in those who discontinue fedratinib.

Clinical advice received by Celgene and the ERG state that, in the absence of alternative effective treatments, patients who are relapsed/refractory to ruxolitinib rarely discontinue ruxolitinib in current UK clinical practice

If fedratinib is recommended in the CDF, data can be captured on a number of outcomes which would complement the FREEDOM-2 trial data and help to resolve the following:

1. Prior treatment
2. The response outcomes in UK clinical practice
3. The discontinuation rate in UK clinical practice
4. Composition of BAT in those who discontinue fedratinib.

The cost-effectiveness model was primarily informed by analyses derived from the JAKARTA-2 trial. JAKARTA-2 was a Phase II, open-label, single-arm study of 97 patients previously treated with ruxolitinib and with intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis. In 2013, the Food and Drug Administration placed fedratinib under a clinical hold due to reported cases of Wernicke's encephalopathy. Consequently, clinical trials investigating fedratinib, such as JAKARTA-2, were terminated prematurely and patients were mandated to stop treatment.

There were numerous issues raised by the ERG during technical engagement relating to the use of JAKARTA-2 as the primary source of trial data. Many of these issues can be resolved when FREEDOM-2 is completed and data from the study can be used within an adaptation of the existing model in that:

- FREEDOM-2 population is consistent with the existing population of interest
- Phase III trial with higher enrolment than JAKARTA-2
- Two-arm trial with direct comparison between fedratinib and BAT, which will include ruxolitinib
- No obscuring of data due to clinical hold

Although FREEDOM-2 will provide direct data to address uncertainty, the cross-over design will limit some of the longer-term outcomes for the BAT arm.

Table 1 outlines how the majority of the issues raised by the ERG can be resolved by entering the CDF and awaiting completion of the phase III FREEDOM-2 study.

Table 1: Issues raised by the ERG and how these may be resolved

Issues raised	How may uncertainty be resolved, beyond efforts in technical engagement
1. Phase II, single arm study design	FREEDOM-2
2. Comparison of fedratinib to BAT	FREEDOM-2
3. Alignment between comparator and modelled population	FREEDOM-2 & data generation within the CDF
4. Modelling approaches	FREEDOM-2, as more mature data will allow more informed modelling choices
5. Omission of supportive care	FREEDOM-2 & data generation within the CDF
6. Inconsistent assumption between fedratinib and BAT	FREEDOM-2 & data generation within the CDF
7. Assumptions on survival	FREEDOM-2 & data generation within the CDF
8. Modelling of stopping rule	Data generation within the CDF will show how fedratinib is discontinued in clinical practice
9. Costs of ruxolitinib	FREEDOM-2
10. Reliability of response	FREEDOM-2
11. End of Life Criteria	FREEDOM-2
Key: BAT, best available therapy; CDF, Cancer Drugs Fund	

2. CHMP definition of responders

The JAKARTA-2 study permitted dose escalation from 400mg daily up to 600 mg daily within the first 6 cycles if there was <50% reduction in spleen size by palpation at the end of cycles 2 and 4. The fedratinib dose could be reduced, interrupted, or discontinued in cases of drug toxicity.[11]

A total of 97 subjects were enrolled in the study (intent-to-treat [ITT] population) and the majority of subjects (n=68; 70.1%) did not have dose up-titration within the first 6 cycles. 29 (29.9%) subjects had their dose up-titrated to 500 mg daily (n=20) or 600 mg daily (n=9) within the first 6 cycles.

At the request of the CHMP, analyses were conducted in which patients who responded after their dose was up-titrated within the first 6 cycles were counted as non-responders.

Spleen response

As shown in Table 2, of 97 patients in the ITT population, a total of 30 patients (30.9%) achieved the primary outcome of a spleen response (defined as the proportion of patients with ≥ 35% spleen volume reduction (SVR) from baseline at the end of cycle 6) without last observation carried forward (LOCF).[11]

The spleen response rate (without LOCF) for patients who received a maximum dose of fedratinib 400 mg daily, thus not counting 8 subjects who achieved a spleen response after their dose had been up-titrated to 500 mg daily (n=3) or 600 mg daily (n=5) during the first 6 cycles as non-responders, was 22.7% (n=22/97).[12]

Corresponding figures for patients who had intermediate-2 or high-risk disease at baseline were [redacted]% (all responders) and [redacted]% (patients who responded on maximum 400mg daily without dose up-titration).

The CHMP considered that the exclusion of patients who received doses >400 mg daily (up-titration) provided a conservative estimate for spleen response rate and that the results were still clinically relevant.[13]

Table 2: JAKARTA-2 Spleen response rate (≥35% SVR) at end of cycle 6 (without LOCF)

Spleen response **End of cycle 6**	**ITT population (N=97)**	**All subjects with response while on 400mg maximum dose for first 6 cycles (N=97)**	**All subjects with intermediate-2 and high-risk disease (N=81)**	**Subjects with intermediate-2 and high-risk disease with response while on 400mg maximum dose for first 6 cycles (N=81)**
n (%)	30 (30.9%)	22 (22.7%)	[redacted]	[redacted]
Key: ITT, intention-to-treat; LOCF, last observation carried forward; SVR, spleen volume reduction.				

Symptom response

Of 97 patients in the ITT population, a total of 24 patients (24.7%) achieved a symptom response (defined as the proportion of patients with ≥ 50% reduction in Total Symptom Score (TSS) at end of cycle 6 using the modified MF-SAF).

Twenty (20.6%) patients in the ITT population achieved a symptom response without requiring dose up-titration above 400 mg daily; 4 patients who achieved a symptom response after their dose had been up-titrated to 500 mg daily (n=3) or 600 mg daily (n=1) during the first 6 cycles were considered non-responders in this analysis.

Corresponding figures for patients who had intermediate-2 or high-risk disease at baseline were [redacted]% (all responders) and [redacted]% (patients who responded on maximum 400mg/day without dose up-titration).

Table 3: JAKARTA-2 Symptom response rate (≥50% reduction in TSS) at end of cycle 6

Symptom response **End of cycle 6**	**ITT population (N=97)**	**All subjects with response while on 400mg maximum dose for first 6 cycles (N=97)**	**All subjects with intermediate-2 and high-risk disease (N=81)**	**Subjects with intermediate-2 and high-risk disease with response while on 400mg maximum dose for first 6 cycles (N=81)**
n (%)	24 (24.7%)	20 (20.6%)	[redacted]	[redacted]
Key: ITT, intention-to-treat; LOCF, last observation carried forward; TSS, total symptom score				

Spleen or symptom response

'Spleen or symptom' response classifies a patient as a responder for meeting either criterion, and is the definition used in the base case analysis submitted to NICE. In the ITT population (N=97), a total of [redacted] patients ([redacted]%) achieved a spleen or symptom response.

Of the patients who did not require dose up-titration above 400 mg daily, [redacted] patients ([redacted]%) in the ITT population achieved a spleen or symptom response.

Corresponding figures for patients who had intermediate-2 or high-risk disease at baseline were [redacted]% (all responders) and [redacted]% (patients who responded on maximum 400mg daily without dose up-titration).

Table 4: JAKARTA-2 Spleen or Symptom response rate at end of cycle 6

Symptom response **End of cycle 6**	**ITT population (N=97)**	**All subjects with response while on 400mg maximum dose for first 6 cycles (N=97)**	**All subjects with intermediate-2 and high-risk disease (N=81)**	**Subjects with intermediate-2 and high-risk disease with response while on 400mg maximum dose for first 6 cycles (N=81)**
n (%)	[redacted]	[redacted]	[redacted]	[redacted]
Key: ITT, intention-to-treat; LOCF, last observation carried forward				

Celgene would like to highlight that most patients in JAKARTA-2 study achieved some degree of reduction in spleen volume and/or reduction in total symptom score as illustrated by the waterfall plots in the original submission. Patients who continue BAT can have a worsening of both outcomes, even when continuing ruxolitinib.

3. Cost-effectiveness model

Following ERG feedback and new data availability, the cost-effectiveness model has evolved over time. Therefore, a brief overview of model iterations is provided to help the reader understand what has changed since the original submission.

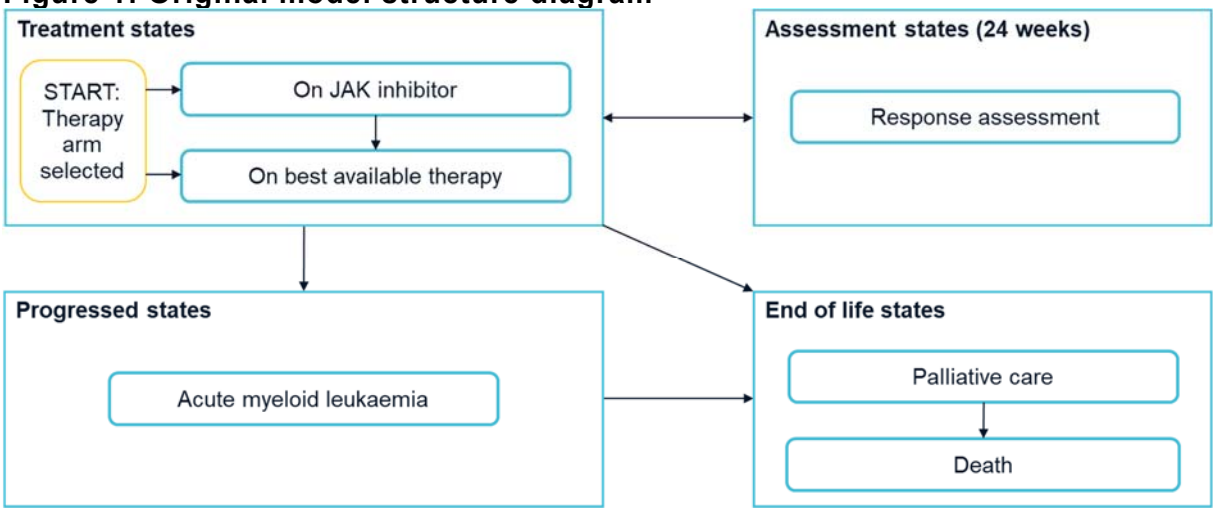
Original submission model

The cost-effectiveness analysis for fedratinib is a discrete event simulation (DES) model built in Microsoft Excel®. In a DES model, patient pathways can be estimated for individuals by sampling directly from time-to-event curves.

The model was primarily informed by the JAKARTA-2 trial, which, as detailed previously, was a Phase II, open-label, single-arm study. The ERG highlighted various concerns with the JAKARTA-2 study, particularly around the methodological limitations of a single-arm study design for assessing clinical effectiveness; to reiterate, these concerns would be resolved by entering the CDF and awaiting completion of the phase III FREEDOM-2 study (see Table 1).

As justified in Celgene's original submission, the base case definition of response is 'Spleen or symptom response', which is a ≥ 35% reduction in spleen volume or ≥ 50% reduction in total symptom score (TSS) from baseline at 24 weeks.

Figure 1: Original model structure diagram



Key: JAK, Janus kinase inhibitor.

Initial clarification questions prompted the following model changes:

- Minor correction to time-in-state output.
- Addition of fedratinib as an optional component of BAT.
- Addition of utility regression outputs that excluded gender as a covariate.
- Addition of option to use HMRN 2020 data for BAT overall survival.
- Addition of option to use Mehra data for BAT overall survival.
- Addition of option to use duration of response data estimated from week 24, rather than data that was originally estimated from 'time of response'.
- Addition of option to cost continuous thiamine supplementation if required.

Technical engagement model

Based on feedback from the ERG[1, 2] the cost-effectiveness model was further updated (Figure 2). These updates included:

Structural changes

- Addition of supportive care health state, with worsening utility over time.
- Replacement of acute myeloid leukemia (AML) health state with implementation as an adverse event.
- Replacement of palliative care health state with one-off end-of-life cost.
- Replacement of instantaneous response assessment health state with assignment of response at the beginning of the simulation.

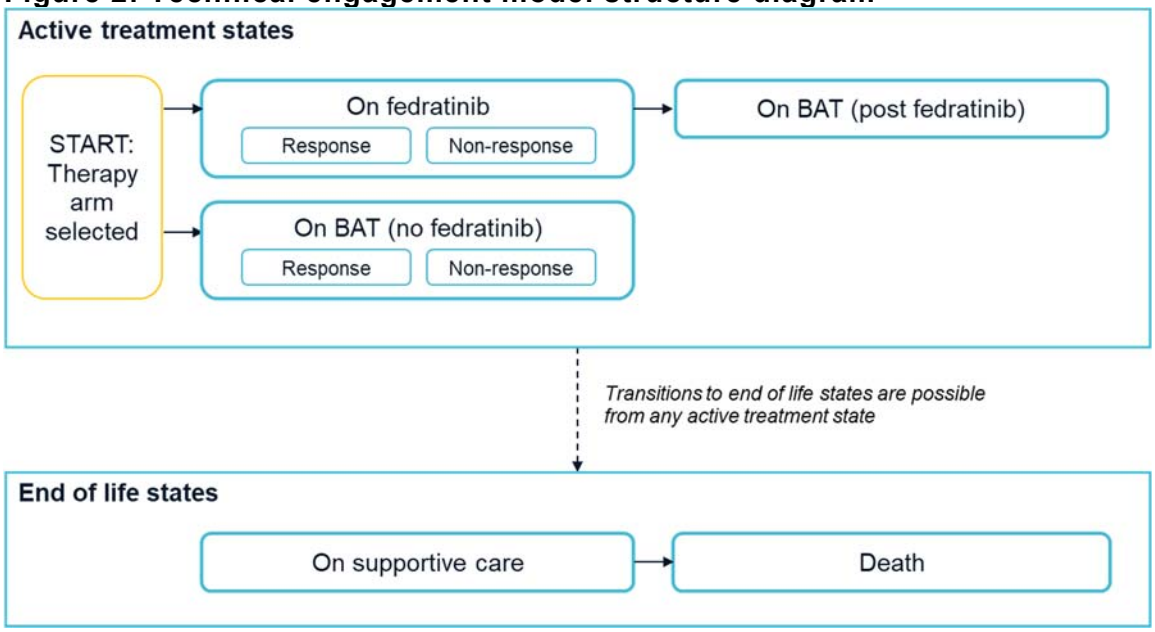
Time-to-event changes

- Addition of functionality for overall survival (OS) to be estimated separately for responders and non-responders, for face validity.
- Addition of functionality for fedratinib OS to be estimated from time of treatment discontinuation (TTD), to resolve concerns regarding face validity of the stopping rule and the independence of OS and TTD.
- Sampling of OS and TTD with a consistent random number (if estimated from model entry) to avoid under-estimation of time on treatment.
- Specification of post-fedratinib transitions (% to BAT, % to supportive care, % remaining time alive in supportive care).
- Removal of duration of response since the data and resultant extrapolations were limited and not reflective of 'spleen or symptom' response.

Data changes

- Addition of TTD curve for BAT to allow patients to discontinue to supportive care (HMRN 2020)
- Addition of chart data evidence for overall survival in patients relapsed, refractory or intolerant to ruxolitinib.
- Addition of SIMPLIFY-2 baseline platelet count distributions as options to support ruxolitinib costing.
- Addition of option to ensure BAT non-responders do not experience a positive increment in utility.
- Replaced response risk differences with odds ratios; and an 'average BAT response' scenario was modelled directly (outside of Excel).
- Simulated 10,000 patients instead of 1,000.

Figure 2: Technical engagement model structure diagram



Key: BAT, best available therapy.

Addendum model

The addendum model builds on the technical engagement model and uses the same structure. Updates were primarily made to inputs that were stratified by response, since the split of responders and non-responders changed following the CHMP opinion. In the model, the new data were entered to replace the ITT population inputs, so that analyses can be compared between original efficacy inputs and new efficacy inputs for the intermediate-2 and high-risk population. Further changes are detailed below:

Corrections

- A correction to the code used to calculate supportive care quality-adjusted life years (QALYs) was made following the ERG identifying an unexpected ratio of QALYs to life years (LYs) in this health state.
- It is acknowledged that inhibiting survival curves from crossing beyond certain time points has limitations. The associated code was updated in efforts to improve the approach.

Functionality updates

- In response to requests during technical engagement, an override switch was added to explore assumptions of equal OS and time on treatment (by using consistent outputs with the fedratinib arm).
- The adjustment to stop fedratinib and BAT OS curves from crossing in the model was altered to follow fedratinib from the point the curves meet, to avoid the long-term plateau in OS associated with Schain 2019 extrapolations.

Input data updates

Response

The CHMP definition of response was not to include any patients who responded following dose up-titration to receive more than 400mg daily of fedratinib, and that the patients who had been up-titrated and then responded were to be classed as 'non-responders'.

Given the number of inputs and assumptions that were informed by the JAKARTA-2 response, the change in the definition of response according to the CHMP required changes to many of the analyses and inputs used within the model. The data updates that were performed were the following:

1. **Response Indirect treatment Comparisons (ITCs)** – These were performed using consistent methods described in the original NICE submission, with updated proportions of responders.
2. **Utility analysis** – This was performed using consistent methods described in the original NICE submission, with updated proportions of responders.
3. **Time to treatment discontinuation** and **overall survival** analysis – Parametric fits were fitted separately to the responders and non-responders according to the new CHMP definition of response.

The aforementioned updates are discussed in turn below.

1. Response ITCs

The 'spleen or symptom' ITC was performed between the available BAT and fedratinib trial data with the endpoint of patients having experienced either SVR or TSS. The base case analysis used within the model was the matching adjusted indirect comparison (MAIC) using SIMPLIFY-2 data for the BAT arm and the Eastern Cooperative Oncology Group Performance Status (ECOG PS) and Dynamic International Prognostic Scoring System (DIPSS) scores for adjustment.

Per the methodology detailed in technical engagement response form, relative treatment effects were applied as odds ratios, and the maximum and minimum BAT response scenarios have been produced using the new CHMP definition of response.

These are used to produce an average treatment effect (Table 5). The active fedratinib response percentage for the SVR or TSS endpoint was taken from the JAKARTA-2 Int-2/High-risk base case population (Table 6) and the response adjustment percentages were applied to produce the active response proportions for each therapy arm within the model (Table 7).

Table 5: Application of response adjustments in base case (CHMP response definition)

Treatment	Odds ratio	Lower bound	Upper bound
BAT (minimum response)	[redacted]	[redacted]	[redacted]
BAT (maximum response)	[redacted]	[redacted]	[redacted]
BAT (average response)	[redacted]	[redacted]	[redacted]

Table 6: JAKARTA-2 Int-2/High-risk spleen or symptom response (CHMP response definition)

Treatment	n	N	%	Source
Fedratinib - JAKARTA-2 ITT - intermediate-2 + high risk	[redacted]	81	[redacted]	JAKARTA-2 post-hoc analysis, intermediate-2 or high risk (post-EMA)

Table 7: Base case response probabilities at 24 weeks (CHMP response definition)

Treatment	Active Probability
Fedratinib	[redacted]
Best available therapy	[redacted]

2. Utility analysis

The resultant utilities applied in the model are shown in Table 8 and Table 9.

Table 8: MF-8D utilities as applied in the model

Utility	Implementation	Female	Male
Baseline	Baseline value	[redacted]	[redacted]
JAK response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
JAK non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
Key: BAT, best available therapy; JAK, Janus kinase; MF-8D, myelofibrosis 8 dimensions.			

Table 9: EORTC-8D utilities as applied in the model

Utility	Implementation	Female	Male
Baseline	Baseline value	[redacted]	[redacted]
JAK response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
JAK non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]
BAT non-response	Change from baseline, starting after 4 weeks in state	[redacted]	[redacted]

Key: BAT, best available therapy; EORTC-8D; European Organisation for Research and Treatment of Cancer 8 dimensions; JAK, Janus kinase.

3. Time to treatment discontinuation and overall survival analysis

As detailed in the technical engagement response form, the base case is to model TTD from week 0 split by responders and non-responders; OS for fedratinib is then modelled from the point of discontinuation split by responders and non-responders.

For the time-to-event outcomes, additional analyses were performed to censor events at the point of up-titration and fit parametric curves to the data using this censoring definition. This was conducted as an effort to remove the influence of up-titration beyond 400mg daily on survival and time on treatment, since this up-titration would not be permitted in clinical practice. The option to use both the up-titration censored and non-censored data was made available in the model.

Figure 3: Overall survival KM data for JAKARTA-2 ITT population – up-titration above 400mg daily censored versus non-censored

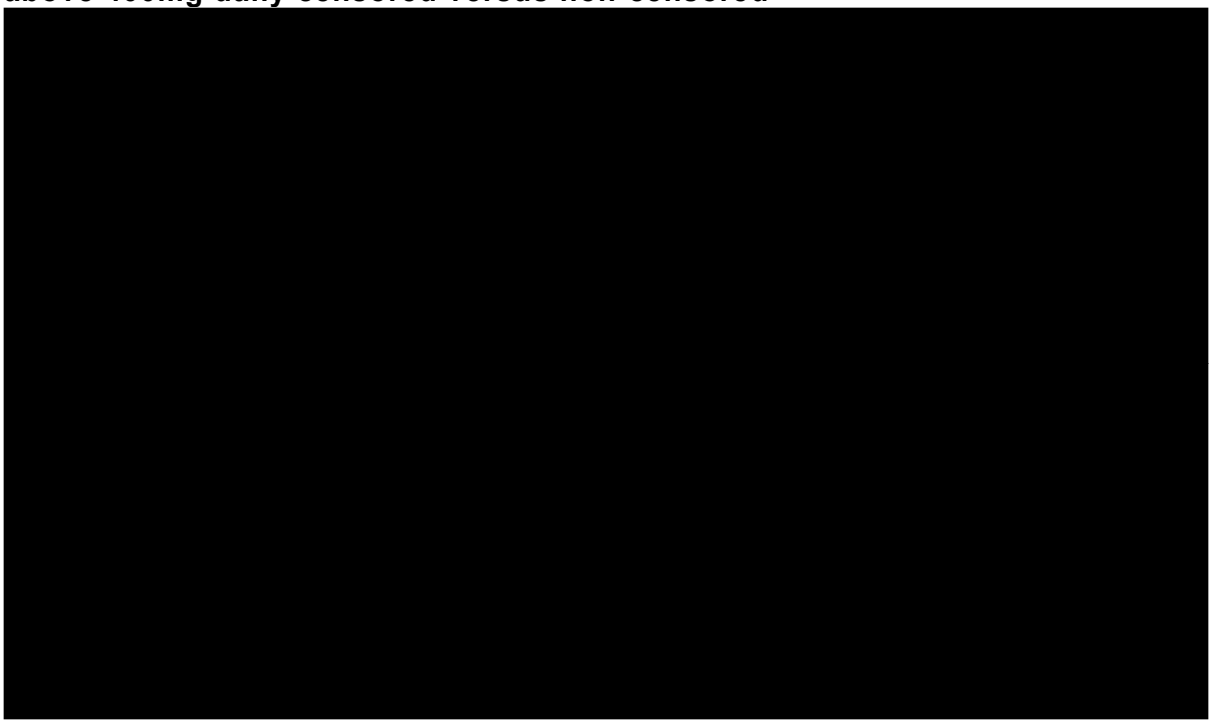


Figure 4: Overall survival KM data for JAKARTA-2 intermediate-2/High-risk population – up-titration above 400mg daily censored versus non-censored

The base case response definition was ‘spleen or symptom’ response. Considering the updated definition of response, the TTD and OS parametric extrapolations used in the model are presented in the below figures.

Figure 5: JAKARTA-2 TTD from 0 weeks – responders (CHMP response definition)

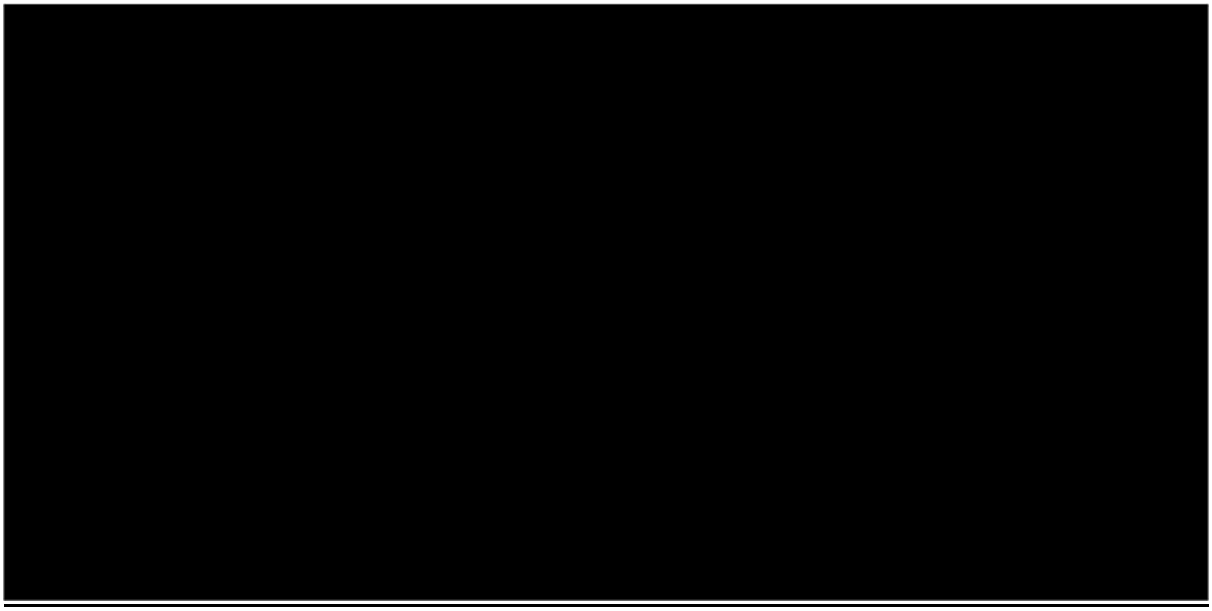


Figure 6: JAKARTA-2 TTD from 0 weeks – non-responders (CHMP response definition)

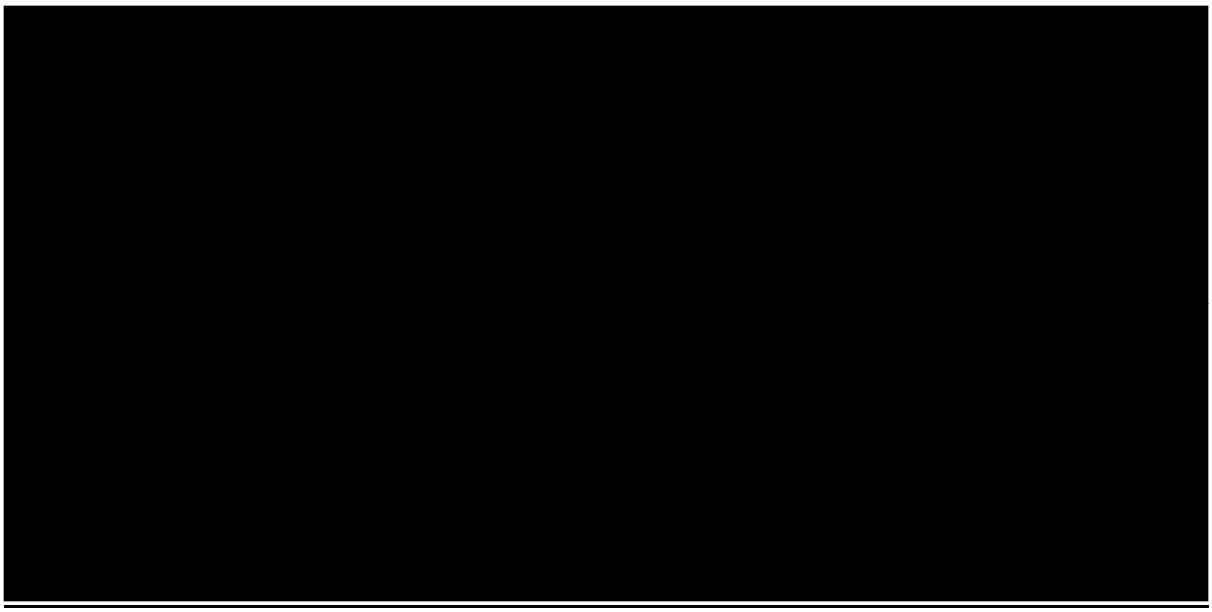


Figure 7: JAKARTA-2 OS from discontinuation – responders (CHMP response definition)

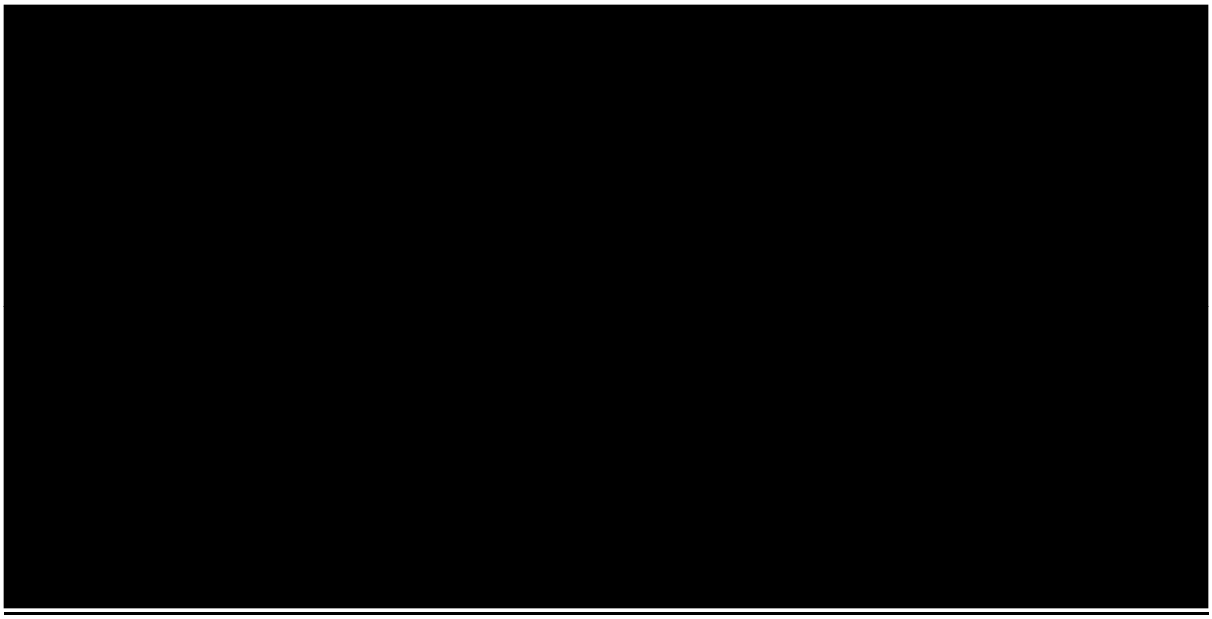
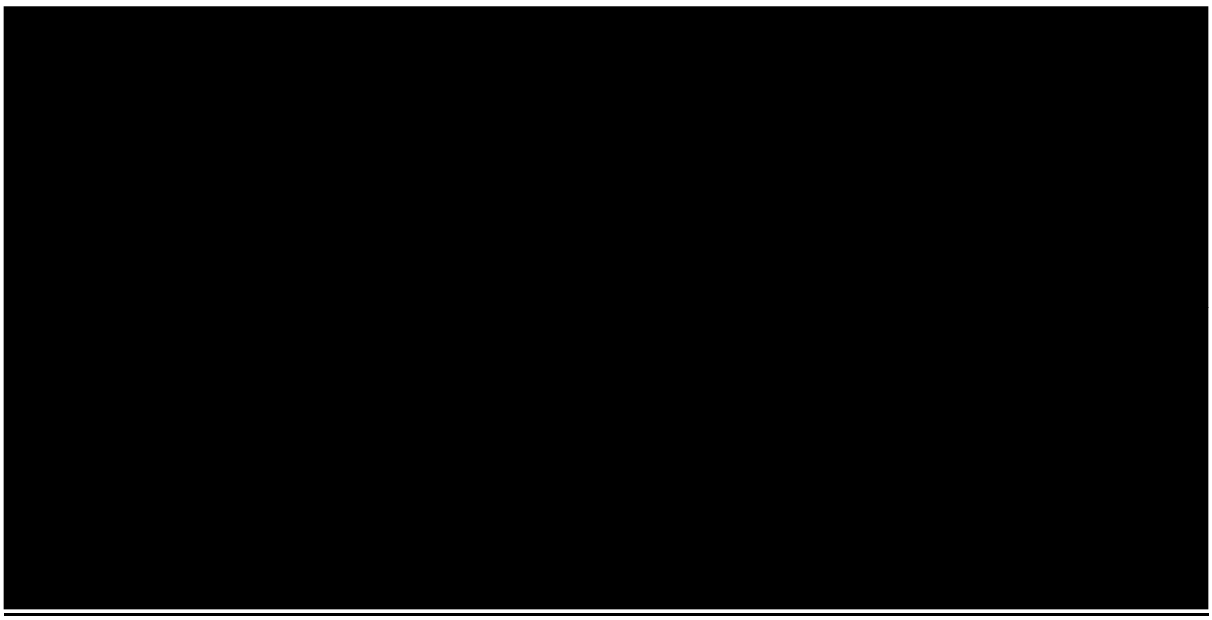


Figure 8: JAKARTA-2 OS from discontinuation – non-responders (CHMP response definition)



Chosen extrapolations and validation

Following the CHMP's amended definition of responders, Celgene sought further advice from a clinical expert on the most appropriate choice of survival extrapolation to choose for the responders and non-responders, at the point of discontinuation. The clinical expert was shown the parametric fits as well as the hazard plots accompanying each fit. The clinical expert stated that it was clinically plausible and reasonable that responders would have a better survival compared to non-responders. For non-responders, the Weibull, Exponential

and Gompertz extrapolations seemed reasonable. When assessing the hazard plots for non-responders, it would be reasonable that the hazard would initially be relatively high and then decrease over time. Therefore, the Weibull was chosen. All the OS curves for responders were deemed similar, except for the Generalised Gamma and Gompertz. Overall, the Weibull was chosen, as the expected hazard would initially be relatively low and would then increase over time. The Weibull was also the most conservative estimate of survival. Further, the Weibull represents a consistent choice of distribution with that made for the BAT extrapolation.

Survival benefit

Celgene reiterates its position that there would be an expected survival benefit with fedratinib in the population of patients who are R/R/I to ruxolitinib and that these patients would be considered end-of-life (EoL). Patients who continue treatment with suboptimal ruxolitinib in this phase would have lost their spleen response, therefore it would be unlikely for them to maintain a survival advantage over fedratinib. The data available from SIMPLIFY-2 shows that there is no survival advantage between patients on BAT (of which 88.5% received ruxolitinib) versus momelotinib. This is most likely driven by there being no significant difference in SVR response between the two groups.

Celgene recognises that there is uncertainty in a survival advantage over BAT due to the single arm design of JAKARTA-2. However, there is supportive evidence as referenced in the original submission to suggest a relationship between SVR and survival. [3, 14-16]

Figure 9: Model output overall survival

Dose intensities

A relative dose intensity (RDI) input for fedratinib is included in the model to accurately capture the impact of dose reductions or missed doses over time on treatment costs.

The actual dose intensity for the ITT population in JAKARTA-2 was [redacted] mg/week, which is equivalent to an average daily dose of [redacted]mg. To assess the actual dose intensity in the absence of dose up-titration beyond 400mg daily, the RDI was calculated by splitting those who were up-titrated and those who were not.

Of the intermediate-2 and high-risk population in JAKARTA-2, for the [redacted] patients where there was no up-titration beyond 400mg daily in the first six cycles, an average of [redacted] mg/week was received. This equates to an average daily dose of [redacted]mg or a RDI of [redacted]%.

For the [redacted] patients who received a dose more than 400mg fedratinib in the first six cycles, it was assumed that these patients would receive the maximum of 2800mg/week in clinical practice.

Therefore, the RDI for the intermediate-2 and high-risk patients (n=81) is calculated as follows:

$$\frac{\text{[redacted]}}{2{,}800\ mg}\ per\ week = \text{[redacted]}\%$$

This RDI equates to an average daily dose of [redacted]mg daily.

In addition, as part of the technical engagement process, it was requested that the company add fedratinib to the basket of therapies in the model that patients may receive after fedratinib discontinuation. This is explored in scenario analysis. For this scenario, a RDI of [redacted] ([redacted]mg daily) is assumed, based on the patients for whom there was no up-titration beyond 400mg daily in the first six cycles.

It is uncertain whether fedratinib would continue as part of BAT in UK clinical practice and, if so, what proportion would continue. An attempt to derive a value based on evidence is made here, although Celgene acknowledges this is limited. In JAKARTA and JAKARTA-2, patients permanently discontinued treatment due to treatment-emergent adverse events (TEAEs). The rates of these observed in the clinical trial programme are shown in Table 10.[13]

Table 10: Fedratinib discontinuation in JAKARTA and JAKARTA-2

Fedratinib Discontinuation	JAKARTA	JAKARTA-2
Up to Cycle 6	13.50%	19.60%
Entire treatment period	24.10%	35%*

*calculated based on the ratio observed in JAKARTA

Due to the clinical hold, the true discontinuation rate in JAKARTA-2 is unknown. Therefore, it was assumed that the ratio observed between cycle 6 and the entire treatment period in JAKARTA (less impacted by the clinical hold) would be similar in JAKARTA-2, and a rate of 35% who permanently discontinued fedratinib was derived. It was assumed that these patients would not receive fedratinib, if fedratinib was to be considered as part of BAT. Therefore, it was assumed that 65% of patients could continue fedratinib as part of BAT in a scenario presented by Celgene.

Price of fedratinib

In acknowledgement of the uncertainty in the existing evidence base, a net price of £[redacted] per pack (representing a discount of [redacted] on the expected list price) has been proposed to reduce the cost of fedratinib.

Results

Updated results

A previous a net price of £[redacted] per pack (representing a discount of [redacted] on the expected list price) was submitted as part of the original submission and technical engagement. The outcomes of the model using the revised model inputs are presented in Table 11. The commercial discount on ruxolitinib is important but unknown, and therefore set to 0%. The model results with a new amended net price are presented in Table 12; this is the revised base case. The base case results show that there is an incremental cost of £11,866, which is less than half of the incremental cost using the previous price of fedratinib. The QALY gain between BAT and fedratinib was calculated to be 0.479, yielding an ICER of £24,784.

The disaggregated clinical and economic outcomes by therapy arm and health state are presented in Table 13. Deterministic sensitivity analysis (Figure 10) and scenario analysis (Figure 11) suggest that key model drivers remain the same.

A probabilistic sensitivity analysis (PSA) was conducted in which all parameters were varied simultaneously over 1,000 iterations, by sampling their values from distributions. The results are summarised in Table 14 and are also presented on a cost-effectiveness plane in Figure 12 and as a cost-effectiveness acceptability curve in Figure 13. 91.3% PSA iterations showed a positive QALY gain for fedratinib over BAT; 23.8% iterations reported that fedratinib resulted in a negative incremental cost for fedratinib. The probability of fedratinib being cost-effective is [redacted]% at a willingness-to-pay (WTP) threshold of £30,000, and [redacted]% at a WTP threshold of £50,000.

Table 11: Revised results summary with previous net price

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	2.394	1.357	-	-	-	-	-
Fedratinib	[redacted]	2.912	1.836	26,300	0.518	0.479	54,929	54,929
Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.								

Table 12: Revised results summary with new net price (base case results)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	2.394	1.357	-	-	-	-	-
Fedratinib	[redacted]	2.912	1.836	11,866	0.518	0.479	24,784	24,784
Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.								

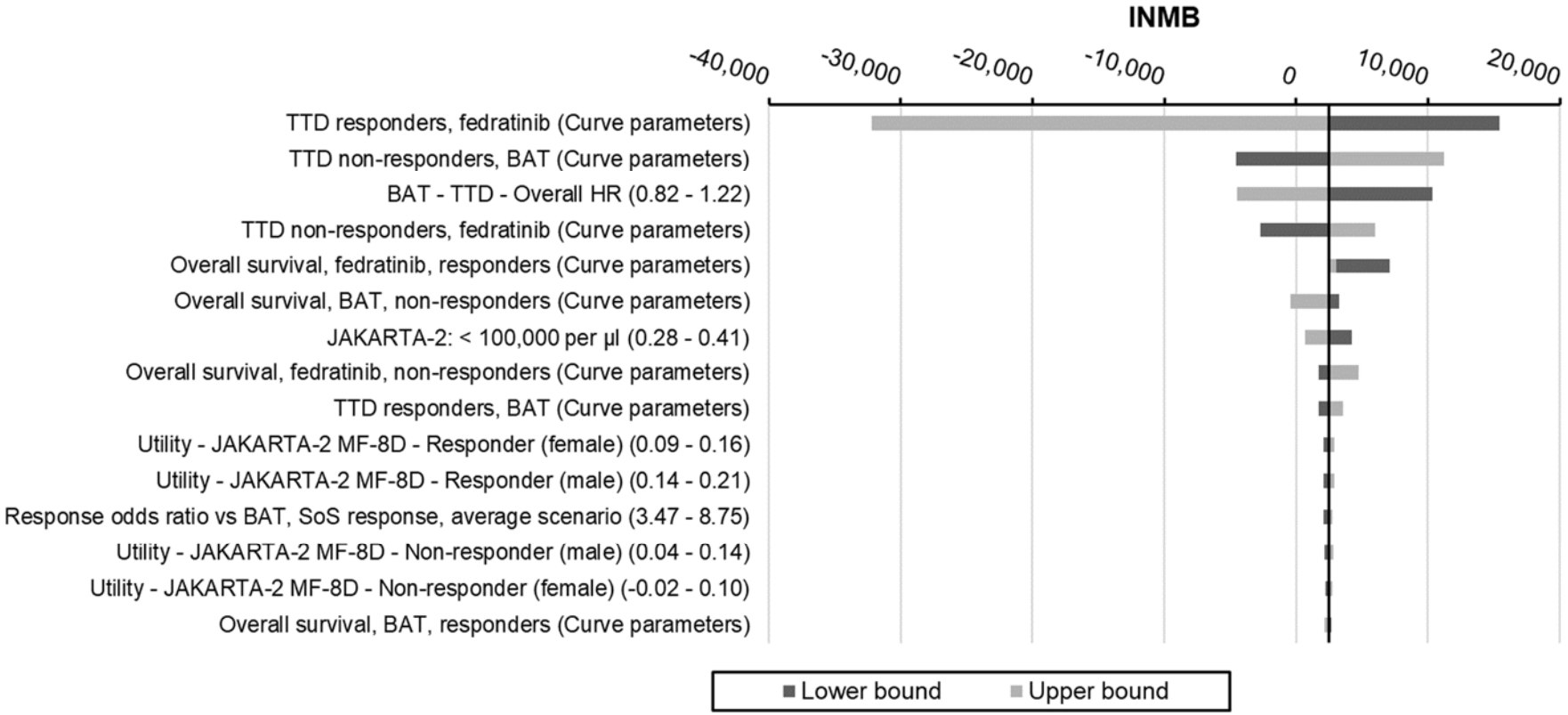
Table 13: Base case disaggregated outcomes

	Treatment arm		Increment	Absolute increment	% absolute increment
	BAT	Fedratinib			
Costs by health state (£)					
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Death (End of life)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	11,866	[redacted]	100%
Costs by category (£)					

	Treatment arm		Increment	Absolute increment	% absolute increment
	BAT	**Fedratinib**			
Acquisition	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adverse events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Resource use	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Thiamine testing and supplementation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
End of life	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	11,866	[redacted]	100%
Life years (LYs)					
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	2.394	2.912	0.518	[redacted]	100%
Median	[redacted]	[redacted]	[redacted]		
Quality-adjusted life years (QALYs)					

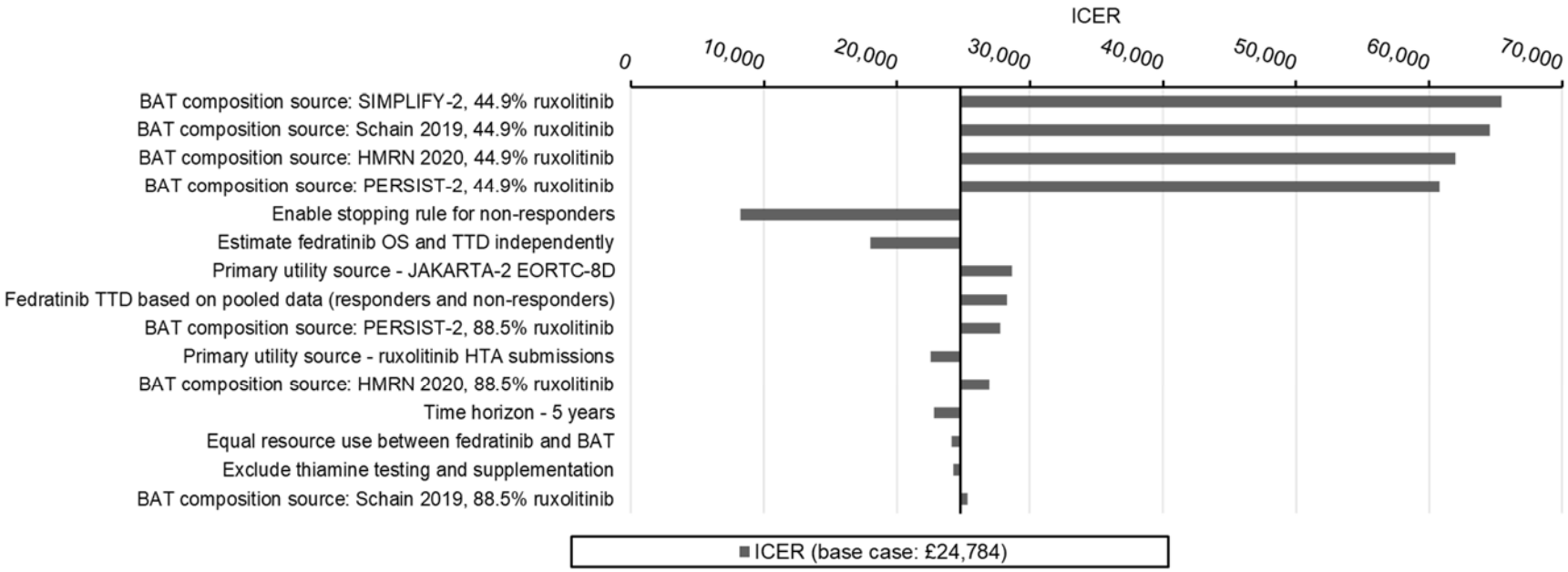
	Treatment arm		Increment	Absolute increment	% absolute increment
	BAT	**Fedratinib**			
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	1.357	1.836	0.479	[redacted]	100%
Median	[redacted]	[redacted]	[redacted]		
Key: BAT, best available therapy; JAK, Janus kinase; JAKi, JAK inhibitor; LYs, life year; QALYs, quality-adjusted life years.					

Figure 10: Results of one-way sensitivity analysis



Key: BAT, best available therapy; HR, hazard ratio; INMB, incremental net monetary benefit; MF-8D, myelofibrosis 8 dimensions; OS, overall survival; SoS, spleen or symptom; TTD, time to treatment discontinuation.
Note: The willingness to pay threshold for INMB calculations was set to £30,000 per QALY.

Figure 11: Results of scenario analysis



Key: BAT, best available therapy; EORTC-8D, preference-based index from the EORTC QLQ-C30; HMRN, Haematological Malignancy Research Network; HTA, health technology appraisal; ICER, incremental net monetary benefit; OS, overall survival; TTD, time to treatment discontinuation.

Table 14: Probabilistic sensitivity analysis results (based on net price)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	3.085	1.445	-	-	-	-	-
Fedratinib	[redacted]	3.983	2.122	19,219	0.897	0.676	28,418	28,418

Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.

Figure 12: Cost-effectiveness plane – Fedratinib vs BAT

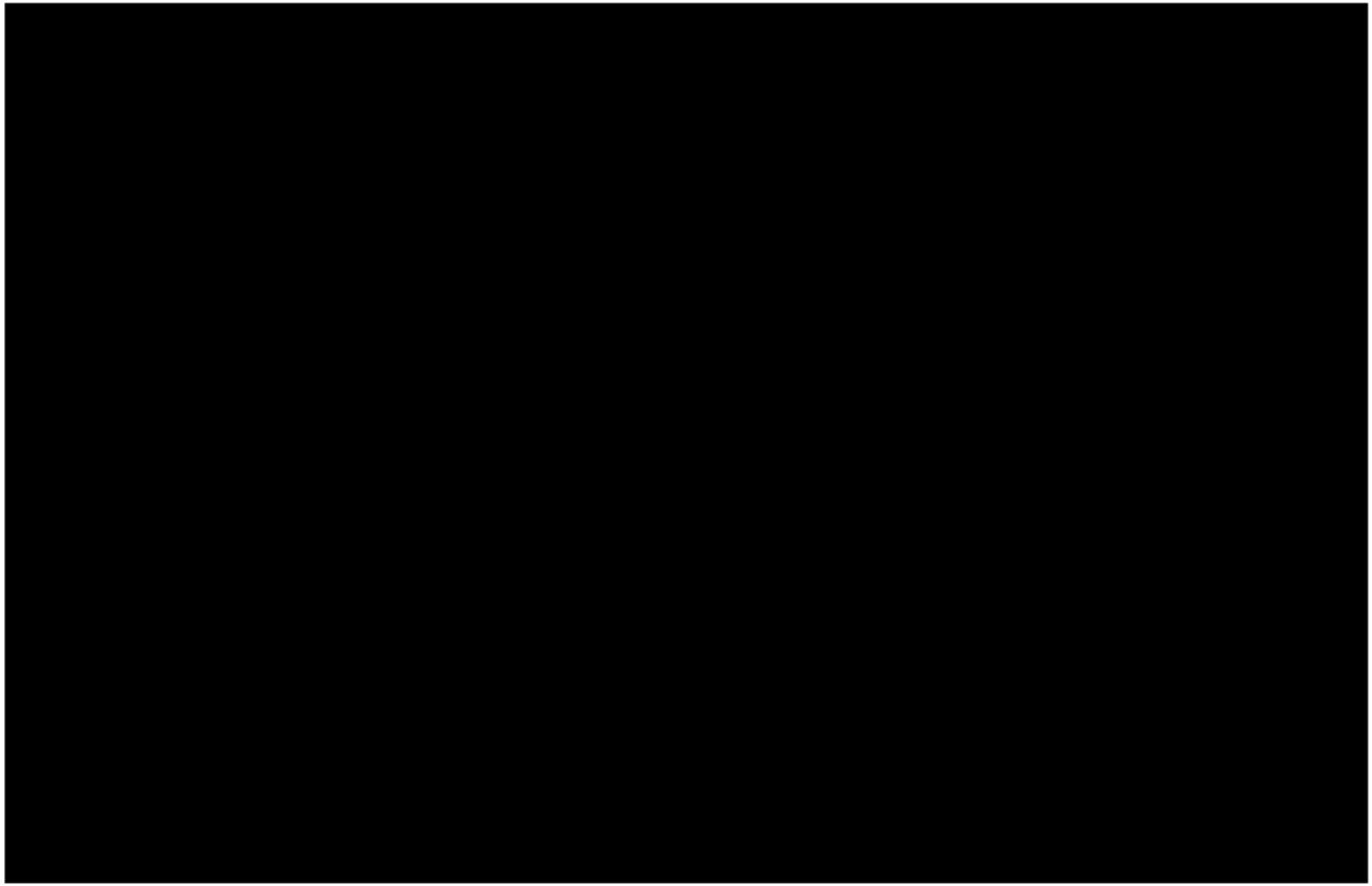


Figure 13: Cost-effectiveness acceptability curve – Fedratinib vs BAT

Key scenarios

As supporting analyses, 3 key scenarios were produced to inform decision making:

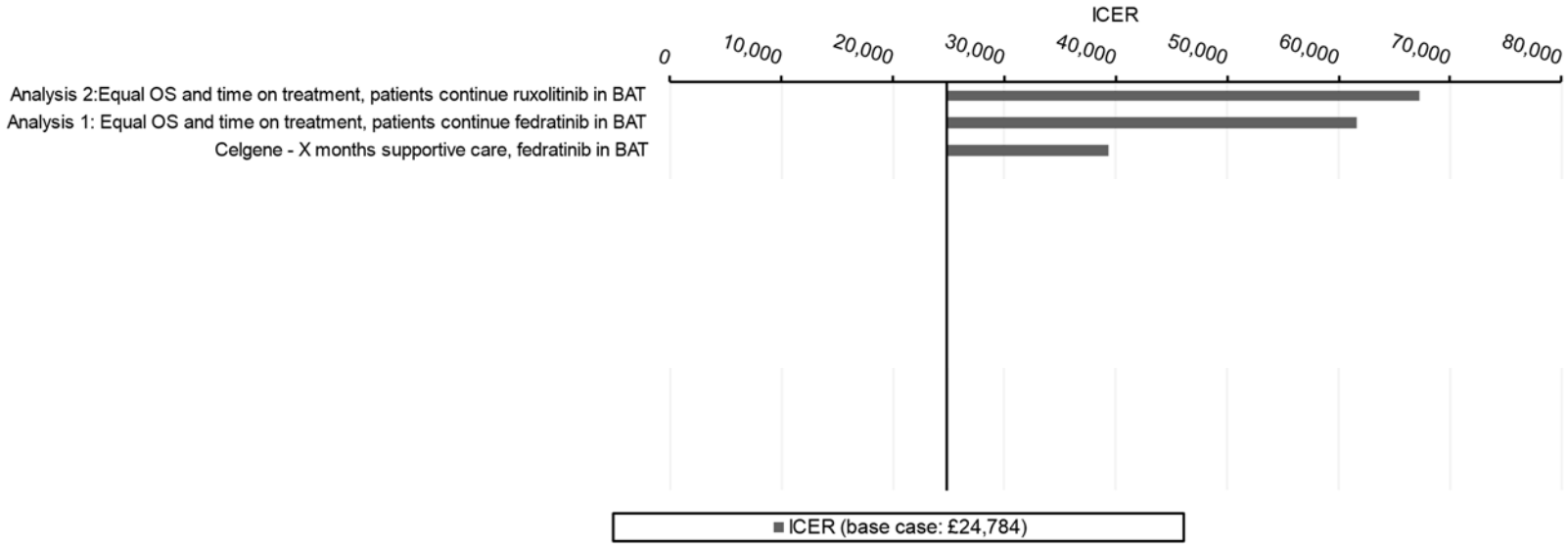
1) 65% continue fedratinib in BAT (see justification in section: *Dose intensities*) and a shorter time of [redacted] months is spent in supportive care for both arms.[17]
2) 'Analysis 1' requested by the ERG. Assumes equal OS and time on treatment between fedratinib and BAT. Patients continue to receive fedratinib in post-fedratinib BAT using the same proportion of patients on ruxolitinib as the BAT arm.
3) 'Analysis 2' requested by the ERG. Assumes equal OS and time on treatment between fedratinib and BAT. Patients receive ruxolitinib in post-fedratinib BAT, using the same proportion of patients on ruxolitinib as the BAT arm.
    a. Celgene would like to note that this request from the ERG is for an unlicensed use of ruxolitinib. Ruxolitinib has not been assessed in the setting after fedratinib nor is it reimbursed for this use.

The detailed results of the scenario analysis are presented in Table 15. A tornado diagram displaying the impact on the ICER is displayed in Figure 14. All three key scenarios produce ICERs higher than a £30,000 WTP threshold, and both the scenarios requested by the ERG produce ICERs higher than a £50,000 WTP threshold. The scenarios where no OS benefit was assumed still accrued QALY gain based on the larger proportion of patients on fedratinib experiencing response. Additionally, the incremental costs for all 3 scenarios are below [redacted].

Table 15: Key scenarios (with net price)

Scenario	Incremental cost	Incremental QALYs	Incremental LYs	ICER
Celgene: Revised base case	11,866	0.479	0.518	24,784
Celgene: [redacted] months supportive care, with fedratinib in BAT	[redacted]	[redacted]	0.518	39,380
Analysis 1: Equal OS and time on treatment, patients continue fedratinib in BAT	[redacted]	[redacted]	0.000	61,582
Analysis 2: Equal OS and time on treatment, patients continue ruxolitinib in BAT	[redacted]	[redacted]	0.000	67,248
Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYs, life years; OS, overall survival; QALYs, quality-adjusted life years.				

Figure 14: Tornado plot of key scenarios



Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; OS, overall survival; X, [redacted].

References

1. School of Health and Related Research (ScHARR) TUoS. Fedratinib for splenomegaly and symptoms in myelofibrosis. A Single Technology Appraisal (ERG report). 15 September 2020 2020. Data on file.
2. Celgene [data on file]. Technical engagement response form Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]. 11 November 2020 2020.
3. Celgene. Myelofibrosis Advisory Board Meeting Report. April 8th 2020 2020.
4. Ragheb M, Harrison CN and McLornan DP. Current and future role of fedratinib in the treatment of myelofibrosis. *Future Oncology*. 2020; 16(6):175-86.
5. Kuykendall AT, Shah S, Talati C, et al. Between a rux and a hard place: evaluating salvage treatment and outcomes in myelofibrosis after ruxolitinib discontinuation. *Ann Hematol*. 2018; 97(3):435-41.
6. National Institute of Care Excellence (NICE). Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. Technology appraisal guidance [TA386]. 2016. (Updated: 23 March 2016) Available at: https://www.nice.org.uk/guidance/TA386/chapter/1-Recommendations. Accessed: 13 November 2019.
7. Palandri F, Breccia M, Bonifacio M, et al. Life after ruxolitinib: Reasons for discontinuation, impact of disease phase, and outcomes in 218 patients with myelofibrosis. *Cancer*. 2019.
8. Schain F, Vago E, Song C, et al. Survival outcomes in myelofibrosis patients treated with ruxolitinib: A population-based cohort study in Sweden and Norway. *Eur J Haematol*. 2019; 103(6):614-9.
9. Jabbour E, Kantarjian HM, Garcia-Manero G, et al. Outcome Of Patients (pts) With Myelofibrosis (MF) After Ruxolutinib (Rux) Therapy. *Blood*. 2013; 122(21):1584-.
10. Mesa RA, Kiladjian J-J, Verstovsek S, et al. Comparison of placebo and best available therapy for the treatment of myelofibrosis in the phase 3 COMFORT studies. *Haematologica*. 2014; 99(2):292-8.
11. Harrison CN, Schaap N, Vannucchi AM, et al. Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the JAKARTA2 study using stringent criteria for ruxolitinib failure. *American Journal of Hematology*. 2020.
12. Celgene. Summary of Product Characteristics (SPC). Inrebic® (fedratinib) 100mg hard capsules. 2020. Available at: TBD. Accessed: TBD.
13. European Medicines Agency (EMA). Assessment report: Inebric. 2020. (Updated: 10 December 2020) Available at: https://www.ema.europa.eu/en/documents/assessment-report/inrebic-epar-public-assessment-report_en.pdf. Accessed: March 2021.
14. Palandri F, Palumbo GA, Bonifacio M, et al. Baseline factors associated with response to ruxolitinib: an independent study on 408 patients with myelofibrosis. *Oncotarget*. 2017; 8(45):79073-86.
15. Vannucchi AM, Kantarjian HM, Kiladjian JJ, et al. A pooled analysis of overall survival in COMFORT-I and COMFORT-II, 2 randomized phase III trials of ruxolitinib for the treatment of myelofibrosis. *Haematologica*. 2015; 100(9):1139-45.
16. Verstovsek S, Kantarjian HM, Estrov Z, et al. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. *Blood*. 2012; 120(6):1202-9.
17. School of Health and Related Research (ScHARR) TUoS. Fedratinib for splenomegaly and symptoms in myelofibrosis. A Single Technology Appraisal. ERG commentary on the company's technical engagement response (Data on File). 27 November 2020 2020.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis ID1501

Clarification questions

August 2020

File name	Version	Contains confidential information	Date
		Yes	

Section A: Clarification on effectiveness data

Literature searching

1. **Priority**. CS B.3.3.3 Overall survival page 116 The August 2018 iteration of the clinical SLR was updated systematically using Embase to identify overall survival evidence for patients after discontinuation of ruxolitinib. Please clarify whether the search was separate and differed from Appendix D.1.1.3 Table 8 (page 11), if the 4,011 publications retrieved (11 reported survival) were from Embase alone as the value differs from the total number of records screened according to Appendix D Figure 1 PRISMA diagram (page 18).

A targeted update to the original systematic review was performed to identify OS evidence for patients with myelofibrosis who had discontinued ruxolitinib. The original systematic review is described in Appendix D.1.1.1, for which Embase, MEDLINE, and the Cochrane Library were searched from database inception until August 2018. The targeted review included searches of Embase until February 2019, and together these sources retrieved 4,011 publications. Further methodology and inclusion/exclusion criteria for the OS targeted review are described in the published poster.[1]

The number of publications retrieved differs from the PRISMA diagram on page 18 as this figure represents the most recent systematic literature review update (Embase, MEDLINE, and Cochrane Library until February 2020).

2. CS Appendix pages 53, 65 and 75 Please provide the full search strategies for CS Appendix G: Published cost-effectiveness studies Appendix H: Health-related quality-of-life studies Appendix I: Cost and healthcare resource identification, measurement and valuation.

The detailed search strategy is presented and uploaded to NICE docs

3. CS Appendix D, Tables 2-4, pages 5-14 Please clarify the reasons for restricting the keyword searching of intervention and comparator terms to title and abstract fields only and excluding drug trade names in the strategies. Please explain the implications of these restrictions on the retrieval of eligible studies from the database searches.

The intervention and comparator terms were searched using a mix of both free-text search terms and Emtree/MeSH terms which removes the possibility of missing relevant studies.

Ruxolitinib background

A1. CS, Page 17: Please can you comment on the statement that, “Of patients treated with ruxolitinib in clinical trials so far, only 28–42% have achieved the primary endpoint of 35% or more spleen volume reduction from baseline.” and its comparison to the response rate observed in JAKARTA-2. The point is to help the ERG put the response rate observed in JAKARTA-2 into context.

The ruxolitinib clinical trial program (COMFORT-I and COMFORT-II) was designed to assess the efficacy of ruxolitinib in a JAK inhibitor naïve population. Comparatively, JAKARTA-2 only included patients that had been treated with ruxolitinib. Becoming relapsed, refractory or intolerant to ruxolitinib is associated with markedly worse clinical outcomes compared with patients who are naïve to JAK inhibitors. As such, the expected observation would be that a lower proportion of patients achieve the primary outcome measure of spleen volume reduction in JAKARTA-2. In this context, the efficacy of fedratinib in JAKARTA-2 was unprecedented as 31% of patients met the primary endpoint, versus a similar proportion of patients with comparatively better prognosis treated with ruxolitinib in COMFORT-I and COMFORT-II.[2-4]

Sample size

A2. Please provide a justification for the sample size calculation and the response rate to rule out for patients treated with standard of care in the study population.

[redacted]

Health-related quality of life from JAKARTA-2

A3. CS, Section B.2.6.4.3, Page 48: Please provide results of an analysis of covariance of EORTC QLQ-C30 and provide 95% confidence intervals for the Intermediate-2/High risk population.

In order to focus on producing responses to questions identified as priority by the ERG we have not provided a response to this question

A4. CS, Section B.2.7, Page 51: Please provide results of an analysis of covariance of EORTC QLQ-C30 and provide 95% confidence intervals for the Intermediate-2/High risk population.

In order to focus on producing responses to questions identified as priority by the ERG we have not provided a response to this question

Adjustments for prognostic factors

A5. Priority. CS, Section B.2.7, Page 51: Please provide results (i.e. coefficients and 95% CIs) of analyses the key outcomes (dichotomised or preferably on a continuous scale) adjusted for the main known or potential prognostic factors (e.g. platelet count, resistant/intolerant) in a single model with continuous variables included as continuous variables (i.e. not dichotomised) for the Intermediate-2/High risk population.

Similar to the multivariable analyses that were performed to assess potential prognostic factors and described in A6, multivariable logistic regression was performed for the SVR and TSS reduction endpoints. These analyses differ to those described in A6 in the following way:

- Models were fitted using data for the Intermediate-2/High risk population.
- Continuous variables were included as continuous variables, where possible.
    - Previously, of the continuous variables, only duration of prior ruxolitinib treatment was dichotomized (to align with how the variable was reported in SIMPLIFY-2). This is now included as a continuous variable for the output below.
- Categorical variables with more than 2 categories were dichotomized.
    - ECOG 0 and 1 were grouped into a single variable, and platelet group was grouped into <100 10^9/L and ≥100 10^9/L

- Resistant/intolerant has additionally been included in the models (the one patient that was neither resistant/intolerant to ruxolitinib was removed from the analyses to avoid problems with complete separation)

The coefficients and 95% confidence intervals are presented in Table 1 and Table 2.

Table 1: Coefficients from the multivariable models for SVR (Intermediate-2 to high-risk population)

Coefficient	Patients with missing data removed			Missing data imputed		
	Estimate	LCI	UCI	Estimate	LCI	UCI
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BMI (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
White race	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG PS 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TSS (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Platelet count ≥100 10^9/L	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen volume (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-PV MF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PMF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intermediate level 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib duration (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hb <10 g/dL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 negative	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 positive	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib resistant	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; Hb, haemoglobin; JAK2, Janus kinase 2; LCI, lower confidence interval; MF, myelofibrosis; PMF, primary myelofibrosis; PV, polycythemia vera; SE, standard error; TSS, total symptom score; UCI, upper confidence interval.

Table 2: Coefficients from the multivariable models for TSS reduction (Intermediate-2 to high-risk population)

Coefficient	Patients with missing data removed			Missing data imputed		
	Estimate	LCI	UCI	Estimate	LCI	UCI
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BMI (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
White race	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Transfusion dependent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG PS 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TSS (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Platelet count >100 10^9/L	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen volume (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-PV MF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PMF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intermediate level 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib duration (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hb <10 g/dL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 negative	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 positive	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib resistant	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; Hb, haemoglobin; JAK2, Janus kinase 2; LCI, lower confidence interval; MF, myelofibrosis; PMF, primary myelofibrosis; PV, polycythemia vera; SE, standard error; TSS, total symptom score; UCI, upper confidence interval.

A6. Priority. CS, Section B.2.93, Page 58: Please provide results of the multivariable analyses that were performed to assess potential prognostic factors.

Multivariable logistic regression was performed for SVR and TSS reduction endpoints. Due to problems with complete separation, transfusion dependence was not included in the multivariable model for SVR. In order to perform forward selection on the multivariable models, missing values for BMI, baseline TSS, and baseline spleen volume were handled in two ways: (1) Subjects with missing information were removed from the analyses, (2) Missing BMI, baseline TSS, and baseline spleen volume were estimated to be the mean of the non-missing values. Missing ECOG PS was estimated to be ECOG PS 1, which was the most prevalent category. Missing JAK2 mutational profile was treated as a separate category, as was recorded in the patient-level data. Forward selection by AIC of the multivariable model for SVR resulted in selection of the following variables:

- ECOG PS (only when missing data is imputed)
- MF subtype (only when subjects with missing data are removed)
- Sex (only when subjects with missing data are removed)
- Age (only when subjects with missing data are removed)
- Baseline TSS (only when subjects with missing data are removed)

Forward selection of the multivariable model for TSS reduction resulted in selection of the following variables:

- Age
- DIPSS
- Baseline spleen volume
- BMI
- ECOG PS (only when missing data is imputed)
- Prior ruxolitinib duration (only when missing data is imputed)

The coefficients, standard errors and p-values from these models are included in Table 3 and Table 4.

Table 3: Coefficients from the multivariable models for SVR (ITT population)

Coefficient	Patients with missing data removed			Missing data imputed		
	Estimate	SE	P-value	Estimate	SE	P-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BMI (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
White race	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG PS 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TSS (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Platelet count ≥100 10^9/L	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen volume (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-PV MF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PMF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intermediate level 1 with symptoms	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intermediate level 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib duration less than 12 weeks	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hb <10 g/dL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 negative	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 positive	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; Hb, haemoglobin; JAK2, Janus kinase 2; MF, myelofibrosis; PMF, primary myelofibrosis; PV, polycythemia vera; SE, standard error; TSS, total symptom score.

Table 4: Coefficients from the multivariable models for TSS reduction (ITT population)

Coefficient	Patients with missing data removed			Missing data imputed		
	Estimate	SE	P-value	Estimate	SE	P-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BMI (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
White race	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Transfusion dependent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG PS 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TSS (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Platelet count >100 10^9/L	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen volume (continuous)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-PV MF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PMF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intermediate level 1 with symptoms	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intermediate level 2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib duration less than 12 weeks	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hb <10 g/dL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 negative	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAK2 positive	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; Hb, haemoglobin; JAK2, Janus kinase 2; MF, myelofibrosis; PMF, primary myelofibrosis; PV, polycythemia vera; SE, standard error; TSS, total symptom score.

A7. CS: The International Prognostic Scoring System (IPSS) define five predictors of survival to determine disease risk in primary MF: age >65 years, haemoglobin (Hgb)

< 10 g/dL, white blood cell (WBC) count > 25 × 109/L, circulating blasts ≥ 1%, and presence of constitutional symptoms. The DIPSS-Plus includes three additional independent prognostic factors: red blood cell (RBC) transfusion dependence, platelet count < 100 × 109/L, and unfavourable karyotype.

i. Please provide summary statistics regarding WBC count in JAKARTA-2. If WBC was not recorded in JAKARTA-2, please provide a reason.
ii. Please provide summary statistics regarding circulating blasts in JAKARTA-2. If circulating blasts were not recorded in JAKARTA-2, please provide a reason.

Summary statistics for WBC count and circulating blasts in JAKARTA-2 are detailed in the table below.

Table 5: Baseline haematology values, JAKARTA-2 (ITT population)

	Fedratinib 400 mg N = 97
WBC count (10^9/L)	
n	[redacted]
Mean (SD)	[redacted]
Median	[redacted]
Min, Max	[redacted]
WBC count > 25 x 109/L, n (%)	
n	[redacted]
Yes	[redacted]
No	[redacted]
Blood blasts > 1%, n (%)	
n	[redacted]
Yes	[redacted]
No	[redacted]
Key: ITT, intent to treat; Max, maximum; Min, minimum; SD, standard deviation; WBC, white blood cell. Source: JAKARTA-2 CSR Table 14.1.8	

iii. Please provide information regarding karyotype in JAKARTA-2.

Karyotype was only collected in JAKARTA-2 for exploratory study purposes (biomarkers). Risk classification in the study was initially performed using IPSS, until this was changed to DIPSS in Global Protocol Amendment 3 (Dated 28 Nov 2012).

Karyotype is not required for IPSS or DIPSS calculation, and DIPSS-Plus was not used in the study.
This aligns with clinical evidence at the time of trial design, which indicated that cytogenetic abnormalities and not the presence of an unfavourable karyotype provided important prognostic information that is not accounted for by the IPSS or other established risk factors.[7]

Indirect treatment comparison

A8. Priority. CS: The ERG believes that the IPSS, DIPSS-Plus and predictors included in other precision models (as well as any other potential prognostic factors and treatment effect modifiers considered relevant by the company) should be included in an appropriate unanchored indirect comparison. Furthermore, to avoid the implication that risk is dichotomous for continuous variables, continuous variables should be included in multivariable models as continuous covariates.

i. Please provide a rationale for why all IPSS and DIPSS-PLUS predictors were not included in the unanchored indirect comparisons.

Please provide results of unanchored indirect comparisons including the IPSS and DIPSS-PLUS predictors and any other potential prognostic factors and treatment effect modifiers. (The ERG notes that absence of evidence is not evidence of absence so that excluding known predictors based on a lack of statistical significance is not appropriate.)

Celgene, a BMS company, would like to highlight that the DIPSS-PLUS was not used in the JAKARTA or SIMPLIFY studies, therefore some of the prognostic factors required for calculation of DIPSS-PLUS are not available.

All available baseline characteristics were considered for inclusion in the matching analyses. Given the relatively small sample size in JAKARTA-2, a strategy for choosing the most important variables in imbalance was used. Variables were included in the matching if they satisfied the following criteria:

- The variable was identified as having clinically meaningful imbalance by an external haematologist
- The variable was also identified as being an important prognostic factor in the JAKARTA-2 study (from either the univariate or multivariable analyses)

Further rationale for not including the IPSS and DIPSS-PLUS predictors in the matching can be found in Table 6.

Table 6: Rationale for why all IPSS and DIPSS-PLUS predictors were not included in the unanchored indirect comparisons

IPSS and DIPSS-PLUS predictors	Rationale for not including the MAIC analyses
Age >65 years	Dichotomised age at baseline (>65, ≤65) was not reported for SIMPLIFY-2. Mean age was reported for SIMPLIFY-2 but was not matched on given the balance across studies (mean [SD] was [redacted] in JAKARTA-2 and 69.4 [7.4] in the BAT arm of SIMPLIFY-2).
Haemoglobin < 10 g/dL	Dichotomised haemoglobin at baseline (< 10 g/dL, ≥10 g/dL) was not reported for SIMPLIFY-2. Mean haemoglobin was reported for SIMPLIFY-2 but was not matched on given the balance across studies (mean [SD] was [redacted] in JAKARTA-2 and 9.5 [1.6] in the BAT arm of SIMPLIFY-2
White blood cell count > 25 × 10^9/L	Not reported for SIMPLIFY-2
Circulating blasts ≥ 1%,	Not reported for SIMPLIFY-2
Presence of constitutional symptoms	Not reported for SIMPLIFY-2
Red blood cell transfusion dependence	Transfusion dependence was reported for SIMPLIFY-2 and was adjusted for in the analyses for spleen volume reduction.
Platelet count < 100 × 10^9/L	Mean platelet count reported SIMPLIFY-2 but was not matched on given balance across studies. Mean platelet count (SD) at baseline was [redacted]) in JAKRTA-2 and 126.5 (95.9) in SIMPLIFY-2.
Unfavourable karyotype	Not reported for either SIMPLIFY-2. In JAKARTA-2 karyotyping was conducted at screening and regular intervals for exploratory (biomarker) purposes.

Additional MAIC analyses were performed to include all possible IPSS/DIPSS-PLUS predictors in the matching. The matching variables were: ECOG PS, DIPSS, transfusion dependence and mean age (and standard deviation). For other variables of interest which were not part of either IPSS or DIPSS, we were unable to match on mean haemoglobin and mean platelet count as baseline values for these variables were not available in the JAKARTA-2 patient-level data (only grouped variables). After matching, the weighted aggregate baseline characteristics for JAKARTA-2 patients were the same as in SIMPLIFY-2, however, the effective sample size was reduced to [redacted] (compared with [redacted] when adjustment is made for just ECOG PS, and [redacted] when adjustment is made for ECOG PS and DIPSS) indicating that the weights are highly variable due to a lack of population overlap, and that the

estimates from these analyses may be unstable (see histogram of rescaled weights in Figure 1). The results show a consistent favourable effect for fedratinib, but the low effective sample size means they should be interpreted with caution.

[redacted]

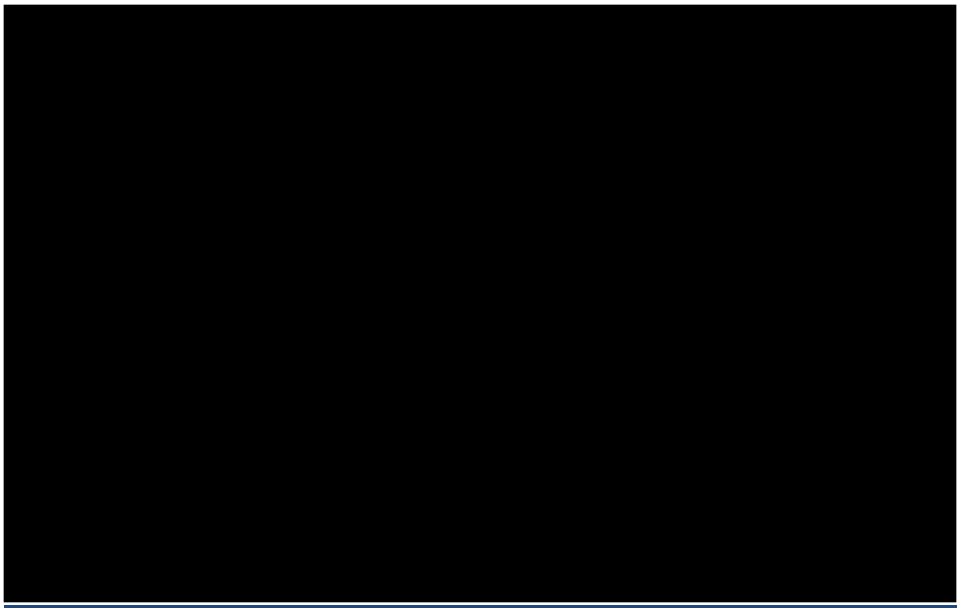


Table 7: Additional MAIC analyses to include all possible IPSS/DIPSS-PLUS predictors in matching

Endpoint	Method	Variables included in adjustment	JAKARTA-2 (400 mg FEDR)	SIMPLIFY-2 (BAT)
SVR	MAIC	ECOG PS DIPSS Transfusion dependence Age	[redacted]% (CI:[redacted])	5.8% (n=3; N=52)
			Risk difference (FEDR versus BAT) [95% CI]: [redacted]	
TSS reduction	MAIC	ECOG PS DIPSS Transfusion dependence Age	[redacted]% (CI: [redacted])	5.9% (n=3; N=51)
			Risk difference (FEDR versus BAT) [95% CI]: [redacted]	

A9. Priority. CS: Please confirm that the indirect comparisons have been conducted on the usual linear predictor scales used for evidence synthesis of the outcomes.

For the MAIC analyses, a risk difference was calculated by:

1. Simulating the SIMPLIFY-2 BAT data based on the number of reported responders and non-responders
2. Combining the simulated comparator data with the JAKARTA-2 IPD
3. Fitting a binomial model with logit link to the combined data that has treatment as a covariate and includes the weights (simulated comparator subjects were assigned a weight of 1)
4. Finally, the proportion of comparator responders predicted from the model was subtracted from the proportion of fedratinib responders, also predicted from the model

To account for the fact that weights were estimated rather than fixed and known, a bootstrap estimator was used to calculate the CI as follows:

1. Fedratinib-treated subjects were sampled with replacement (a bootstrap dataset)
2. For each bootstrap dataset, a set of weights was derived
3. For each bootstrap dataset and corresponding set of weights, a proportion of fedratinib-treated responders was obtained
4. For each bootstrapped sample, the risk difference was calculated by subtracting a simulated comparator proportion (by assuming a normal distribution) from the proportion of fedratinib treated responders

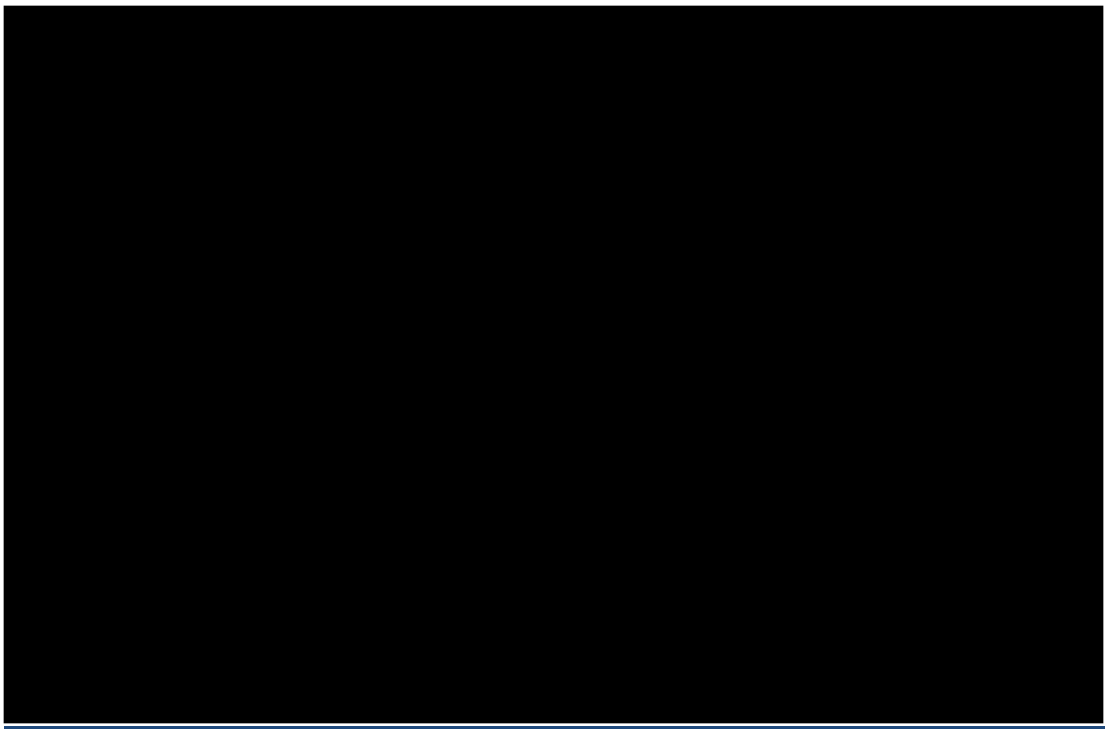
This procedure was repeated 10,000 times to obtain a distribution of proportions which was used to calculate the CIs of the risk difference.

A10. Priority. CS: Please provide histograms of the distribution of propensity score weight used in the matching-adjusted indirect comparisons.

Figure 2: Histograms of the rescaled weights (weights applied to the JAKARTA-2 data for comparison of SVR) - ECOG PS and transfusion dependence adjustment

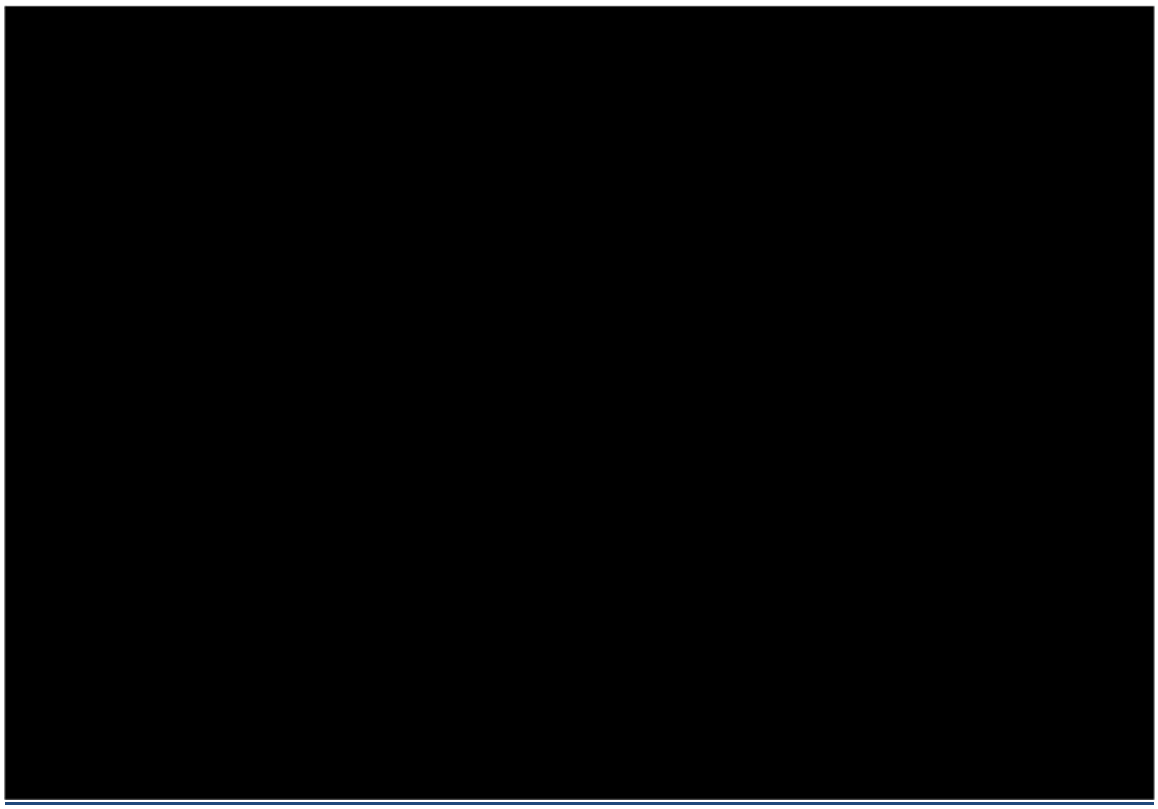
ECOG PS, Eastern Cooperative Oncology Group Performance Status; SVR, spleen volume reduction.

Figure 3: Histograms of the rescaled weights (weights applied to the JAKARTA-2 data for comparison of SVR) - ECOG PS adjustment



ECOG PS, Eastern Cooperative Oncology Group Performance Status; SVR, spleen volume reduction.

Figure 4: Histogram of the rescaled weights (weights applied to the JAKARTA-2 data for comparison of TSS reduction) - DIPSS and ECOG PS adjustment



DIPSS, Dynamic International Prognostic Scoring System; ECOG PS, Eastern Cooperative Oncology Group Performance Status; TSS, total symptom score.

A11. Please use data from JAKARTA to investigate potential treatment effect modifiers in a multivariable model.

It was not considered suitable to use the JAKARTA patient-level data to identify treatment effect modifiers in the post-ruxolitinib setting given that all patients in JAKARTA had not received ruxolitinib and disease is expected to be more severe post-ruxolitinib. Not being able to identify treatment effect modifiers in the post-ruxolitinib setting was highlighted as a limitation of the unanchored MAIC analyses. In the context of ITCs, for patients with no prior ruxolitinib treatment, treatment effect modifiers in the JAKARTA data were explored. Penalized logistic regression models with an interaction term for randomized treatment and each baseline characteristic in Table 8 were fitted to the JAKARTA data for the endpoints of SVR and TSS reduction. A penalized likelihood ratio test was performed to assess the significance of the interaction term. Potential treatment effect modification was identified for p-values < 0.1. The following variables were potential treatment treatment-effect modifiers for SVR:

- JAK-2 status
- Constitutional symptoms

For TSS reduction, no treatment treatment-effect modifiers were identified.

Subgroup analyses for SVR with the corresponding interaction p-values based on a penalized likelihood ratio test are presented in Table 8.

Table 8: Subgroup analyses for SVR[a] with the corresponding interaction p-values based on a penalized likelihood ratio test

Variable (category 1, category 2)	RD[b] (95% CI) category 1	RD[b] (95% CI) category 2	Interaction p-value
Age (≤65, >65)	[redacted]	[redacted]	[redacted]
ECOG PS (≥1, 0)	[redacted]	[redacted]	[redacted]
Race (White, not White)	[redacted]	[redacted]	[redacted]
Sex (female, males)	[redacted]	[redacted]	[redacted]
Weight (≤median, >median)	[redacted]	[redacted]	[redacted]
Haemoglobin (≤10g/dL, >10g/dL)	[redacted]	[redacted]	[redacted]
LDH (≤5 ULN, >5 ULN)	[redacted]	[redacted]	[redacted]
Platelet count (<100x10^9/L, ≥100x10^9/L)	[redacted]	[redacted]	[redacted]
WBC (<25x109/L, ≥25x10^9/L)	[redacted]	[redacted]	[redacted]
Transfusion dependent (no, yes)	[redacted]	[redacted]	[redacted]
Blasts (<1%, ≥1%)	[redacted]	[redacted]	[redacted]
Fibrosis grade (1 or 2, 3)[c]	[redacted]	[redacted]	[redacted]
JAK2 mutation (negative, positive)	[redacted]	[redacted]	[redacted]
Spleen size >10cm (no, yes)	[redacted]	[redacted]	[redacted]
Spleen volume(≤median, >median)	[redacted]	[redacted]	[redacted]
Constitutional symptoms (no, yes)	[redacted]	[redacted]	[redacted]

Notes: a, Endpoint defined as the proportion of patients achieving ≥ 35% spleen volume reduction from baseline to End of Cycle 6 (Week 24); b, Risk difference for fedratinib versus placebo; c, Four patients in JAKARTA had fibrosis grade 0 (one in the fedratinib arm and three in the placebo arm), the patient in the fedratinib arm had a SVR and no patients in the placebo arm had a SVR.

A12. Priority. CS, Section B.2.9.3, Page 61: Please confirm the result of the feasibility of conducting an unanchored indirect comparison with respect to overall survival?

The feasibility of conducting an unanchored indirect comparison was summarised in Appendix L.5.1. The table presents the studies examined for feasibility and the available data for each. For a study to inform an unanchored indirect comparison, it needed to include OS Kaplan Meier data and report sufficient data on baseline characteristics. It was found that none of the studies examined fulfilled these criteria, therefore the analysis could not be conducted.

Response rates from JAKARTA-2

A13. CS, Page 41: The CS states that, "Based on Kaplan–Meier (KM) estimates, only 25% of patients had a duration of response of less than 9.4 months …". Please confirm that this does not include 50 patients who were considered to be non-responders without any duration of response.

Yes, this interpretation is correct.

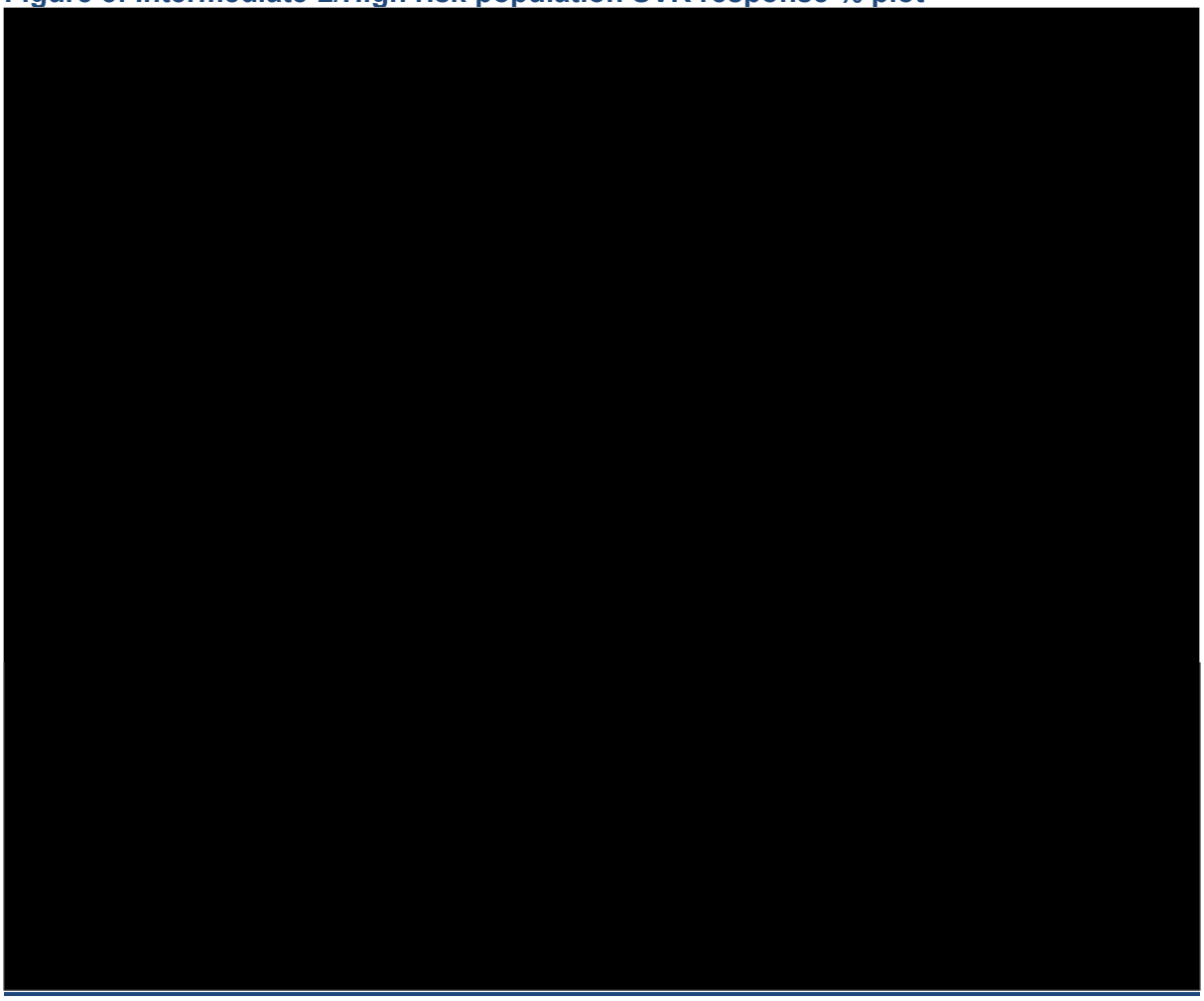
A14. Priority. CS, Section 2.6.3.3, Page 42: Please clarify why the range of percentage changes in EOC3 and EOC6 do not include the median percentage changes within the ranges.

This is a misprint, the upper limits of these ranges should be positive rather than negative, i.e.: 'The median percentage changes in spleen volume were [redacted]% at EOC3 (range: [redacted]) and -38.0% at EOC6 (range: -73, 115).[6, 8]'

A15. CS, Page 42, Figure 8: Please clarify why the number of patients does not correspond to the number in the respective populations, and provide the plot for the Intermediate-2/High risk population.

The number of patients reflects the number with recorded measurements at EOC6.[8] The equivalent graph for the intermediate-2/High risk population is displayed in Figure 5.

Figure 5: Intermediate-2/High risk population SVR response % plot



A16. CS, Section B.3.3.2.4, Page 107: The CS states that "*The number of BAT patients who reach the endpoint is equal to the maximum number of patients experiencing either SVR or TSS response separately – Referred to henceforth as the Minimum BAT response scenario.*

Please confirm that a better minimum response would be the minimum number of patients who experience either outcome.

We disagree with this suggestion. This may be worded confusingly in the documentation. If in an example we know there are 5 SVR responders and 3 TSS responders, then we know there must be at least 5 'SVR or TSS' responders. We believe that should be the minimum, which is conservative. In the maximum response scenario, the number of responders would be 8 (5 + 3), as this assumes that spleen volume responders and symptom responders are distinct.

A17. Priority. CS, Table 39: Please provide the adjusted number of responders and effective sample sizes for the fedratinib arm.

The number of responders and effective sample sizes for both the minimum and maximum scenarios are the same because the scenarios alter the number of responders for the BAT arm and not the fedratinib arm. These values are therefore presented above these scenarios in Table 35 – Table 37 of the CS.

A18. CS, Section B.3.3.2.5:

i. Figure 16 - Please provide a new copy with the time scale labelled on the x-axis.
ii. Figure 16 - Please confirm the number of patients who lost response to treatment i.e. the number of patients with an event.
iii. Figure 16 - Please describe the *a priori* clinically expected shape of the hazard function.
iv. Figure 16 - Please provide parameter estimates, variance-covariance matrix and 95% confidence intervals.

Figure 16 - Please check the exponential model as the ERG thinks there may be a mistake.

i. The time unit on the chart is in Years. A new copy of the chart is added below and has been altered in the model

ii. Of the patients who responded, two had a duration of response event

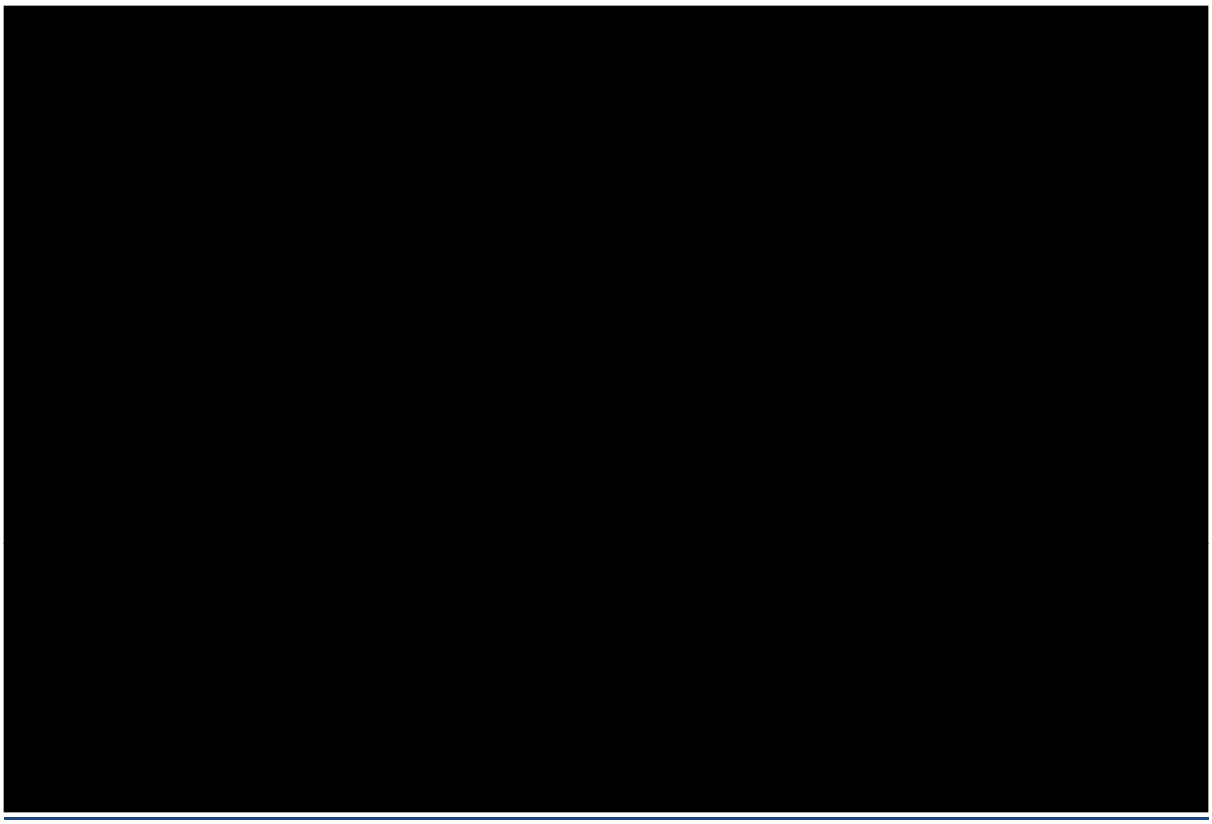
iii. The clinical study design team did not consider the shape of hazard functions *a priory* and clinicians were not consulted on their expectations for duration of response. Clinical advice was sought and indicated an expectation of a decreasing hazard over time. The selected log-normal curve has an initial increase in hazards, followed by a decrease in hazards.

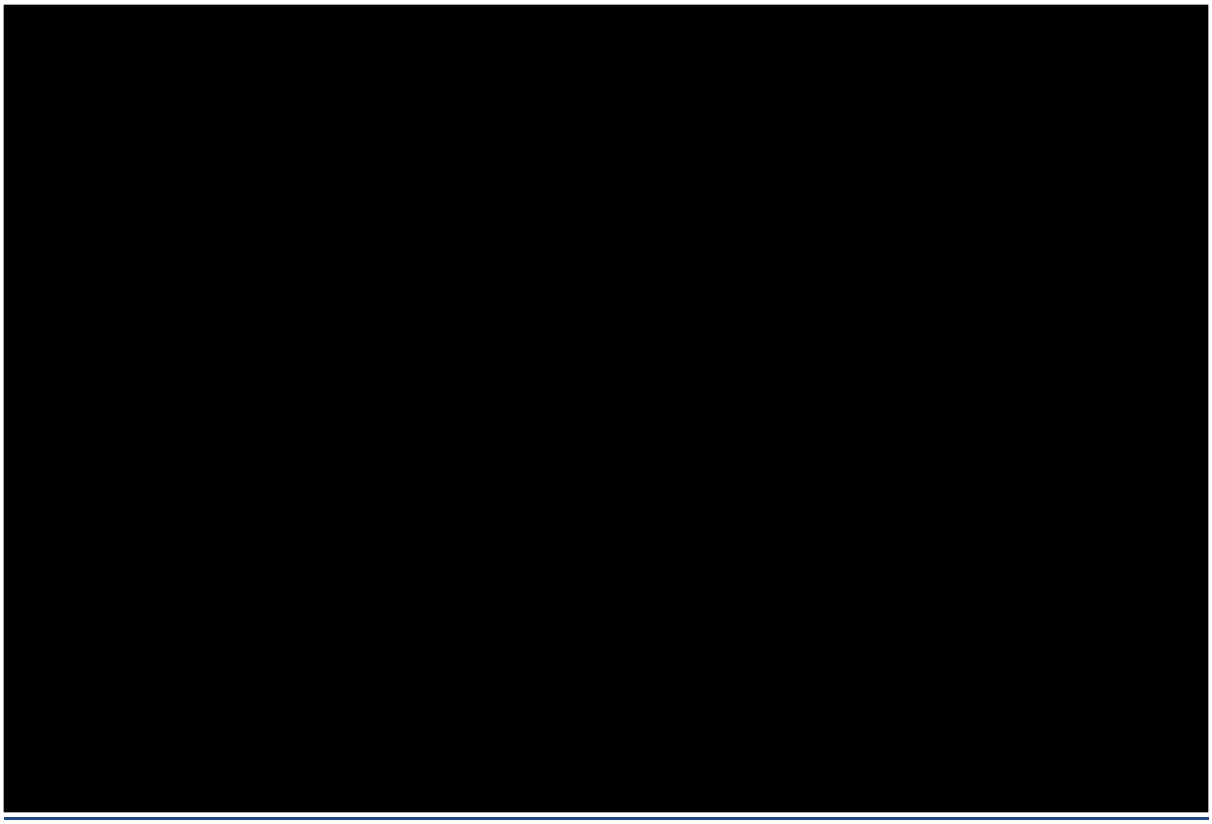
iv. The parameter estimates and variance-covariance matrices are provided in the 'DOR Data' page in the model. The parameters and graphs for the 95% CI are presented below.

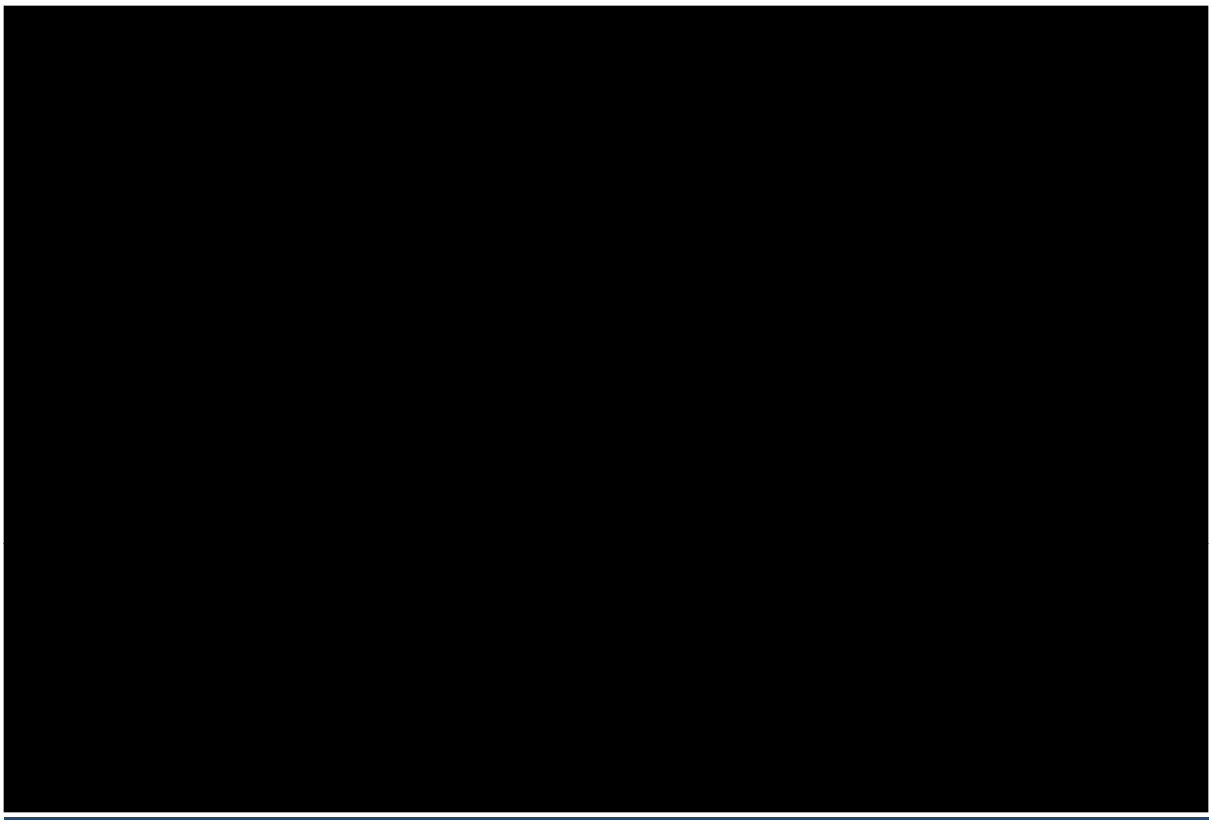
Distribution	Parameter	est	L95.	U95.
Exponential	rate			
Weibull	shape			
Weibull	scale			
Gompertz	shape			
Gompertz	rate			

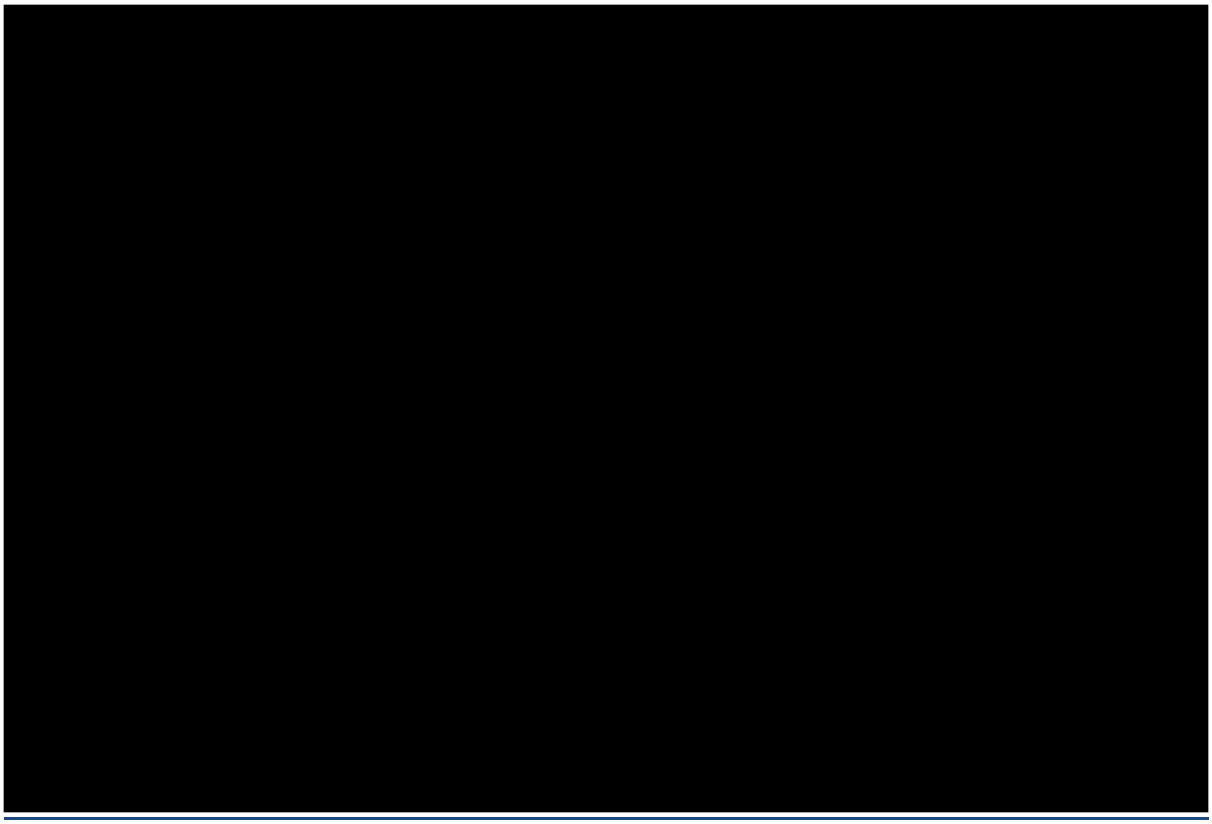
LogLogistic	shape	[redacted]	[redacted]	[redacted]
LogLogistic	scale	[redacted]	[redacted]	[redacted]
LogNormal	meanlog	[redacted]	[redacted]	[redacted]
LogNormal	sdlog	[redacted]	[redacted]	[redacted]

[redacted]









We have double-checked the extrapolated values and found that the exponential rate used in the graph is correct. This extrapolation can be explained by the very

small number of events and the inherent assumption of constant hazards for an exponential distribution.

Overall survival

A19. CS, Section B.2.12, Page 73: Please provide evidence in support of the statement that fedratinib delivers a clinically meaningful survival gain.

In absence of a defined threshold for establishing a clinically meaningful survival gain in myelofibrosis patients, this has to be inferred using the evidence available. Findings that support this conclusion include:

- Published reports have demonstrated a median OS of 13–16 months in myelofibrosis patients following ruxolitinib discontinuation.[10-13] Of ruxolitinib discontinued patients receiving fedratinib in JAKARTA-2, the majority ([redacted]%) were still alive at 12 months suggesting a trend towards prolonged OS compared with previous reports (with an associated improvement in HRQoL)
- In patients that have been treated with ruxolitinib, the economic model predicted [redacted] additional life years with fedratinib compared to BAT. This is equivalent to an additional [redacted] months of survival, almost doubling the OS previously reported for this population.
- Considering that NICE End-of-Life EoL criteria is based on a 3-month extension of life, the [redacted] month estimate suggests that the survival gain provided by fedratinib is significant for patients treated with ruxolitinib

A20. Priority. CS, Section B.2.13, Page 75: The CS states that “The JAKARTA and JAKARTA-2 trials reported similar OS rates for fedratinib 400 mg at 12 months ([redacted]% for JAKARTA-2 and [redacted]% for JAKARTA).” Please confirm in what sense these response rates are considered to be similar given that they imply a hazard ratio of 0.493, and provide an explanation as to why these proportions are so different.

JAKARTA measured the efficacy of fedratinib in a JAK-naïve population, whereas only patients that had received ruxolitinib were included in JAKARTA-2. Given that in clinical practice becoming relapsed, refractory or intolerant to ruxolitinib is associated with a negative impact on symptoms and survival outcomes, the expected observation may be to see a much greater difference in OS between these two studies. A percentage difference of patients alive at 1-year < 10% between the two

studies supports the use of fedratinib as a highly efficacious treatment in the more severe post-ruxolitinib setting.

A21. Priority. CS, Section B.3.3.3:

i. Please describe the *a priori* clinically expected shapes of the hazard functions for each treatment group.
ii. Please provide parameter estimates, variance-covariance matrix and 95% confidence intervals for each model.
iii. The CS, Page 125 states that, "*The Weibull curve was selected in the base case as it provided a better statistical fit to the data.*" Please clarify this statement given the information criterion presented in Table 49.

The six parametric distributions that were used to model overall survival data have restrictive hazard shapes and none are likely to be the true model. For overall survival, please evaluate more flexible models such as fractional polynomials and restricted cubic splines.

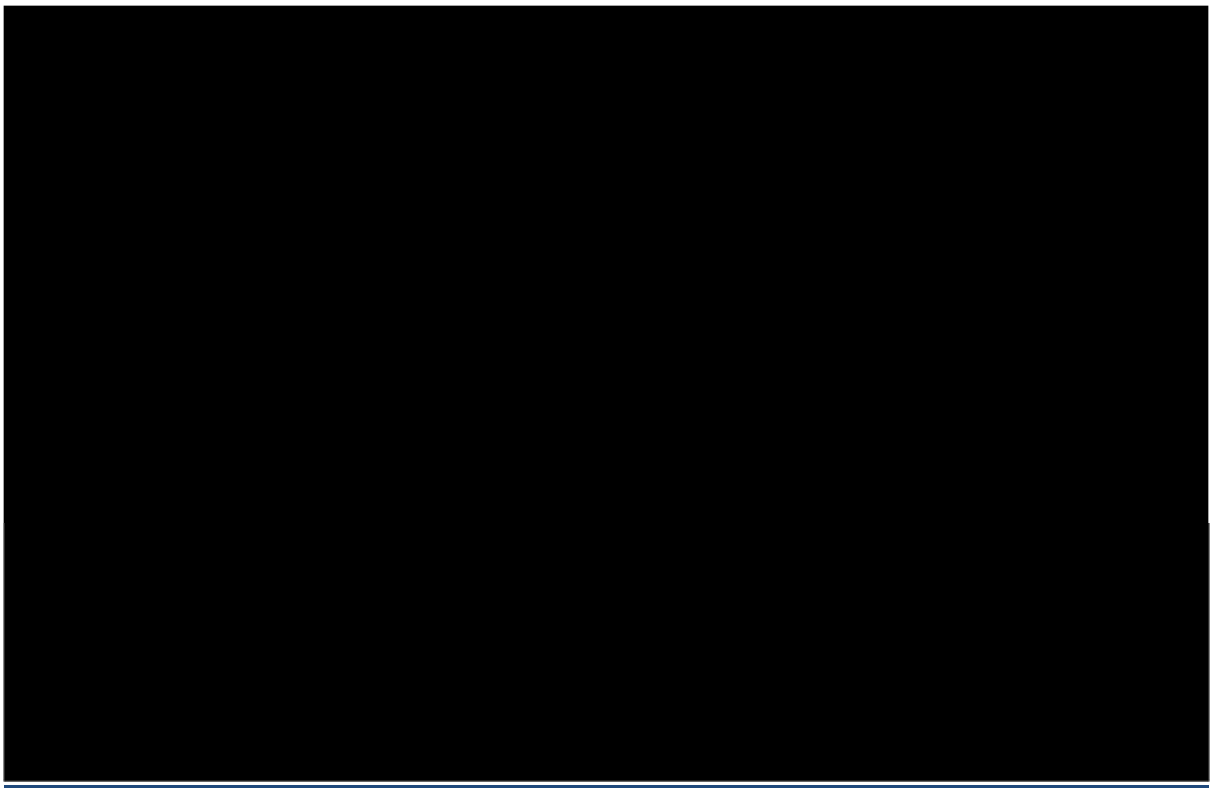
i. The clinical study design team did not consider the shape of hazard functions *a priory*. Clinical advice was sought and indicated an expectation of a decreasing hazard over time for fedratinib OS, whereas the selected curve for fedratinib (Gompertz) shows an increasing hazard.
ii. The parameter estimates and variance-covariance matrices are provided in the 'OS Data' page in the model. The parameters and graphs for the 95% CI are presented below.

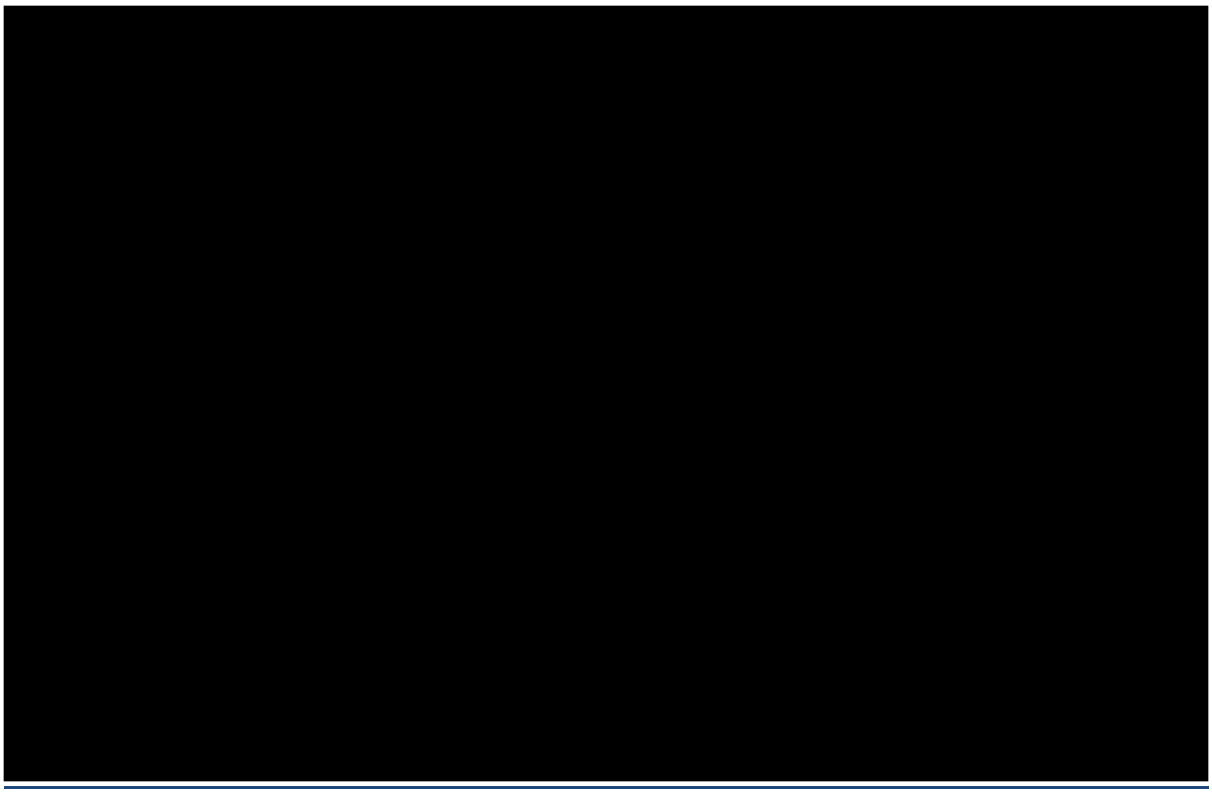
Distribution	Parameter	est	L95.	U95.
Exponential	rate	[redacted]	[redacted]	[redacted]
Weibull	shape	[redacted]	[redacted]	[redacted]
Weibull	scale	[redacted]	[redacted]	[redacted]
Gompertz	shape	[redacted]	[redacted]	[redacted]
Gompertz	rate	[redacted]	[redacted]	[redacted]
LogLogistic	shape	[redacted]	[redacted]	[redacted]
LogLogistic	scale	[redacted]	[redacted]	[redacted]
LogNormal	meanlog	[redacted]	[redacted]	[redacted]
LogNormal	sdlog	[redacted]	[redacted]	[redacted]

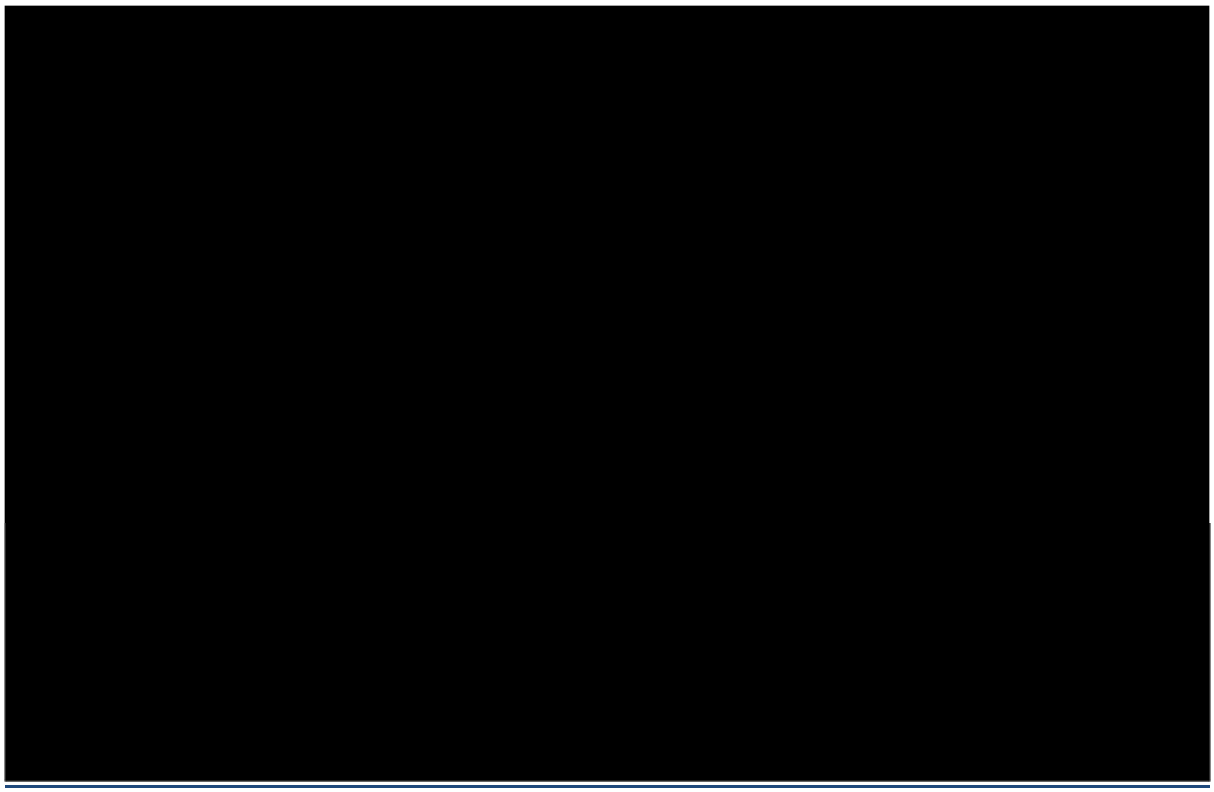
GenGamma	mu	[redacted]	[redacted]	[redacted]
GenGamma	sigma	[redacted]	[redacted]	[redacted]
GenGamma	Q	[redacted]	[redacted]	[redacted]

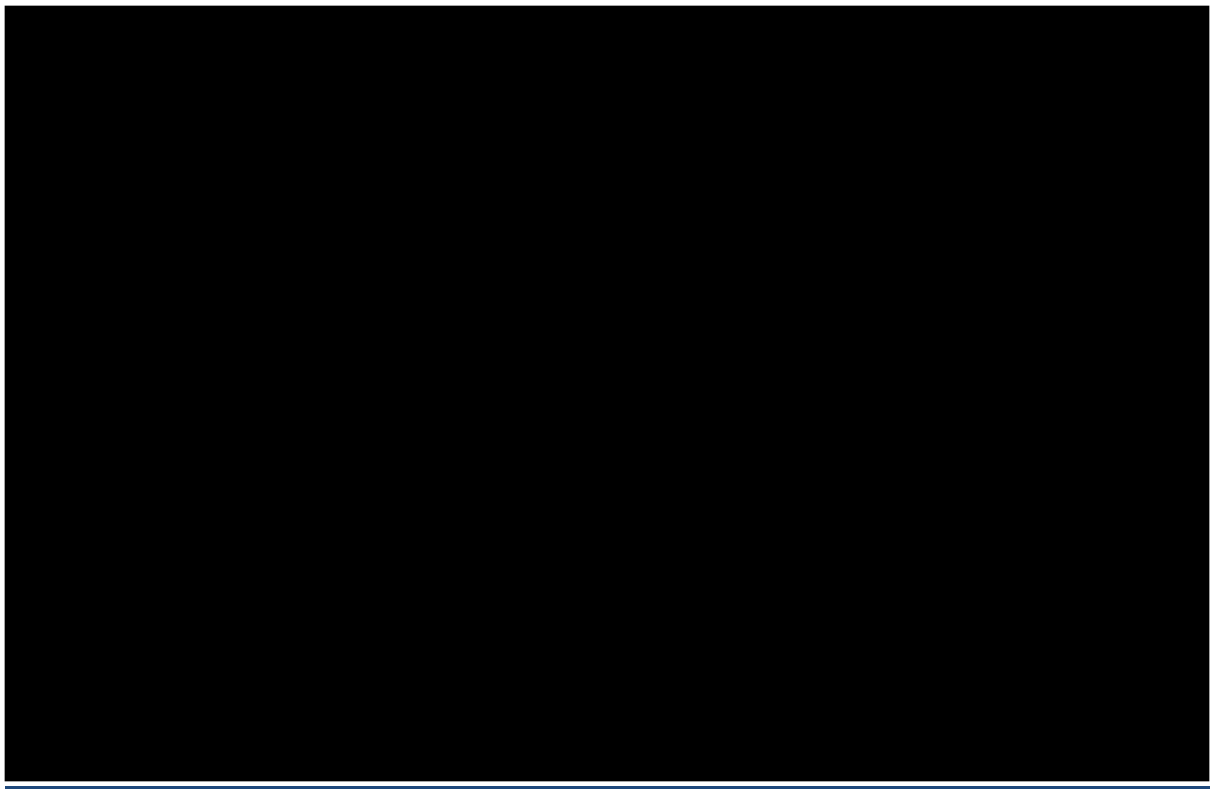
[redacted]

[redacted]









iii. The preceding sentence on page 125 states that the clinical advisory group indicated that the exponential and Weibull curves were the most

representative of UK patients. The Weibull curve was selected as the base case because it was a better statistical fit than the exponential curve

Given the immaturity of the data, the six parametric distributions were appropriate for extrapolating patient overall survival. Using more flexible models is dependent on assumptions regarding a change in hazard rate at one or more specified timepoints. There was not enough data to make these assumptions, therefore the application of these methods in the model would be poorly informed and difficult to justify.

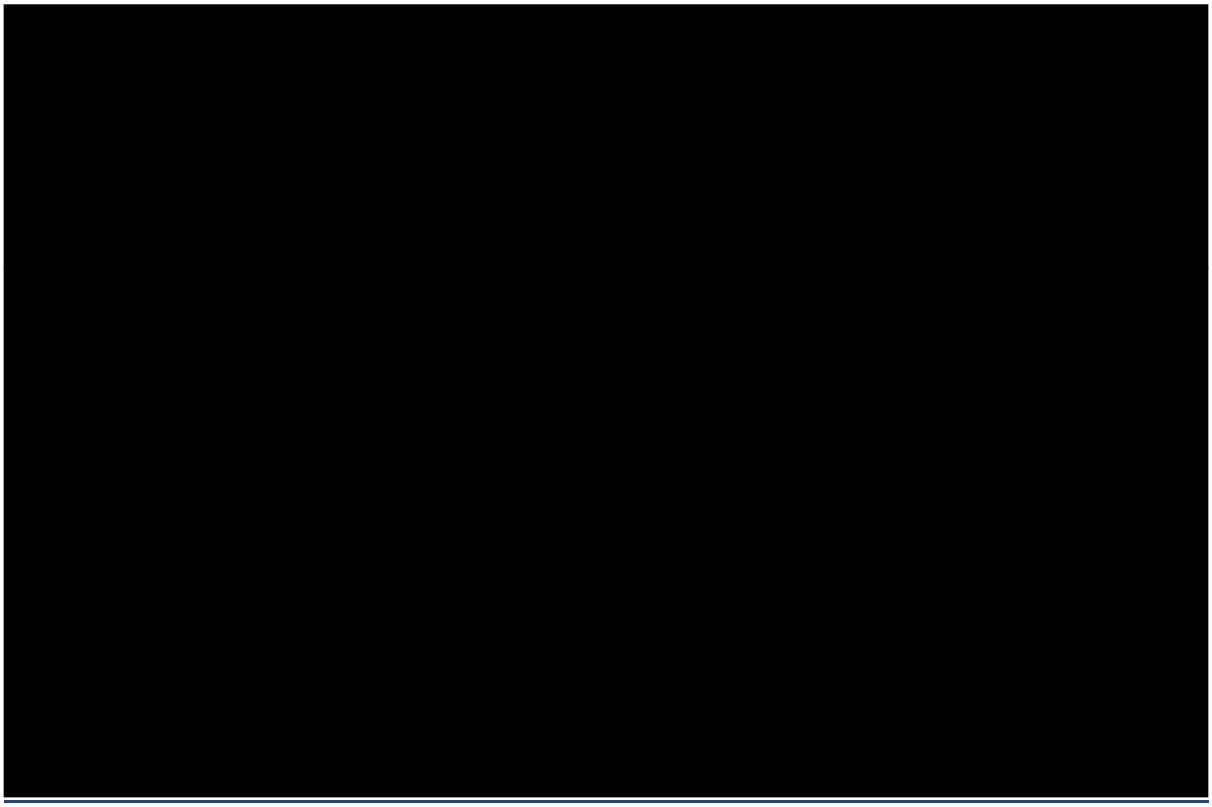
A22. Priority. CS, Page 129, Fedratinib Overall Survival:

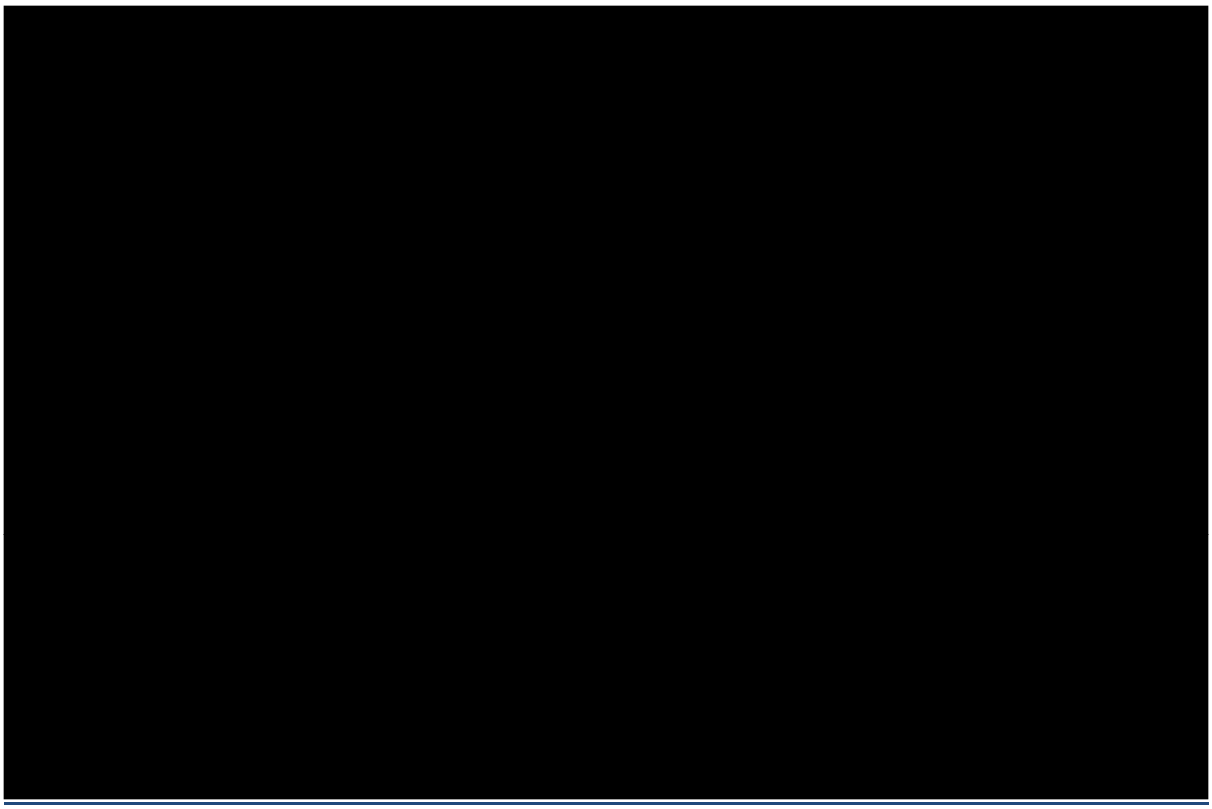
i. Please confirm that the fedratinib overall survival hazard is expected to be increasing over time.
ii. Please provide a plot of the smoothed overall survival empirical hazard with 95% confidence intervals and the overall survival hazard function from the fitted Gompertz distribution.

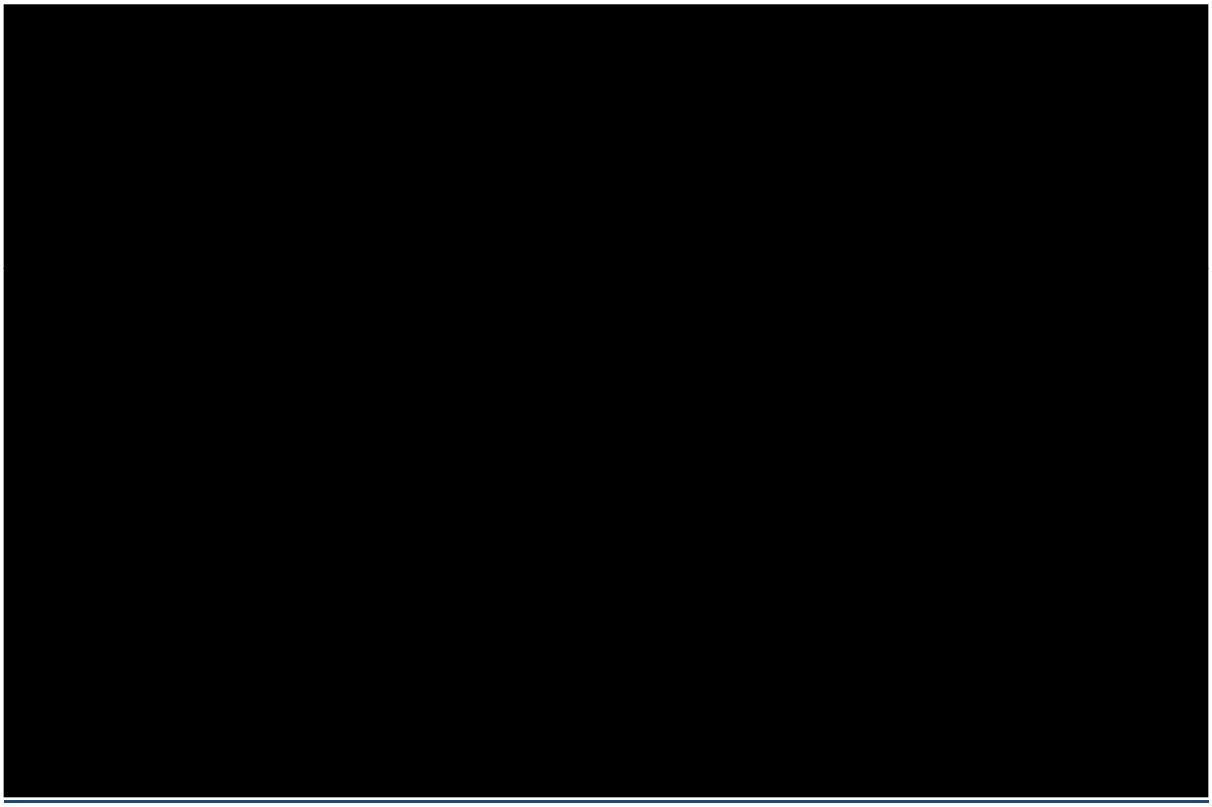
Please provide parameter estimates, variance-covariance matrix and 95% confidence intervals for each model.

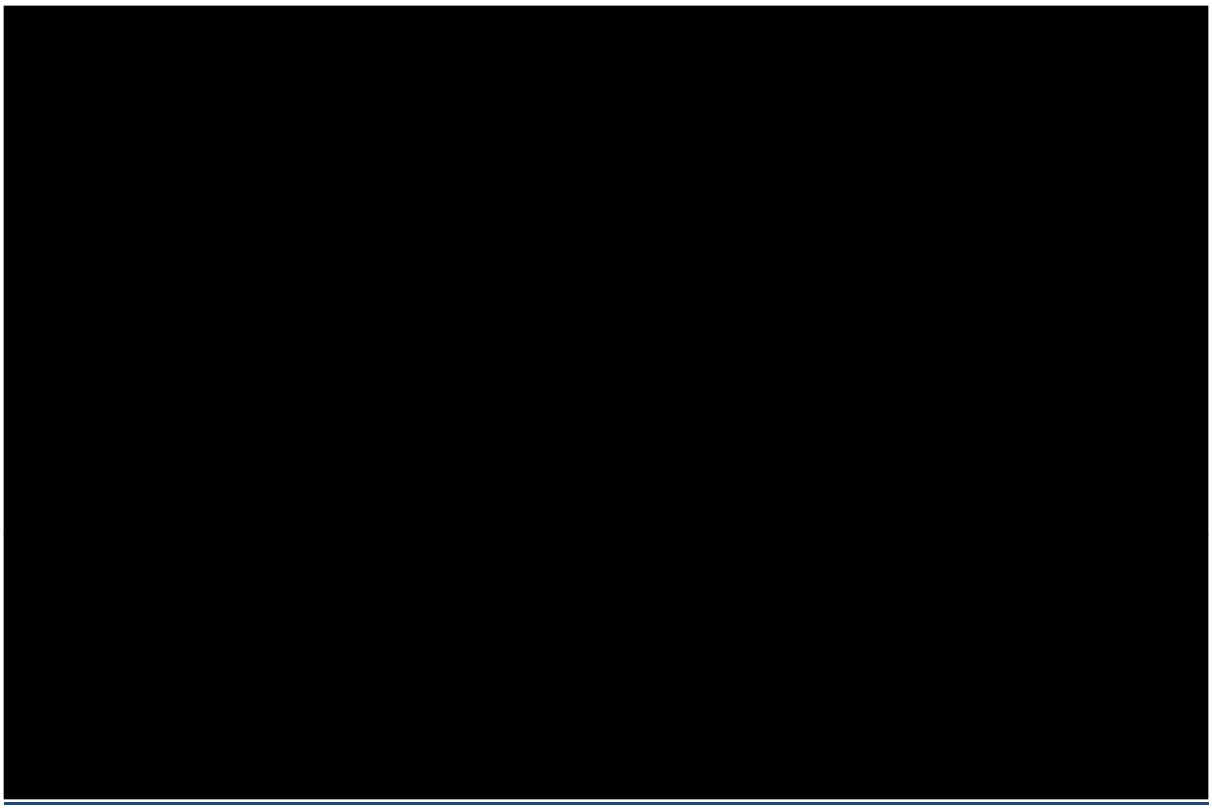
i. The *a priori* expectations of clinicians for OS hazards over time were not collected. For OS, only the survival estimates over time were collected *a priori*. The KM data appeared to show an increase in hazard over time (see graphs below).
As described above, clinical advice was sought and indicated an expectation of a decreasing hazard over time for fedratinib OS, whereas the selected curve for fedratinib (Gompertz) shows an increasing hazard.

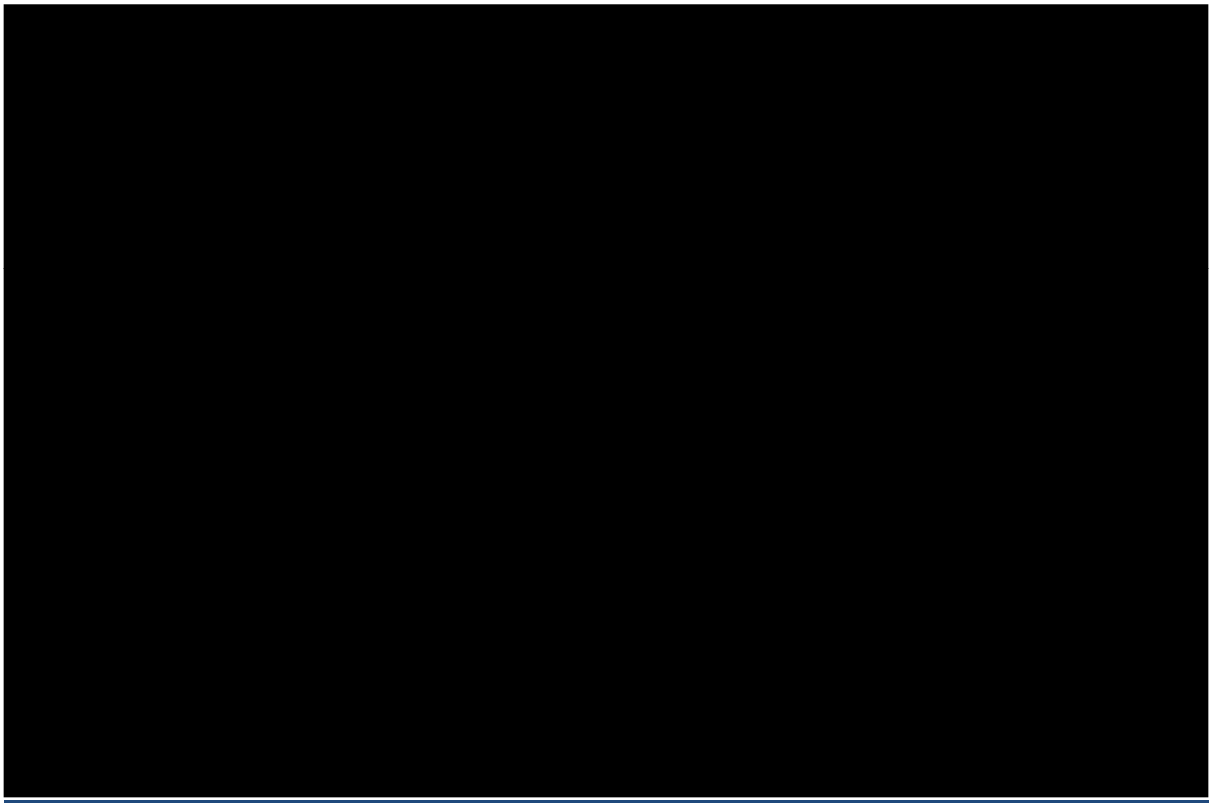
ii. Smoothed overall survival empirical hazards with 95% confidence intervals from the OS data for intermediate-2 and high-risk patients are presented on each chart below, alongside the hazards and 95% confidence intervals predicted by the parametric models. In the charts below, the black lines are from the data (using muhaz function and a bootstrapped for 95% CI) and blue lines are from the parametric models.











Appendices

A23. Priority. Appendix M:

i. Please clarify the parameter values expressed as percentages when beta distributions are used.
ii. Please confirm other parameters expressed using beta distributions. For example, for “Disutility per event: Abdominal pain” the parameter are 0.09 and 0.13, which has mean 0.409, whereas the table suggests a mean of 0.11.
iii. Please clarify why several uncertain parameters are not given uncertain distributions.

Please confirm the parameter values used in normal distributions and why the means do not correspond to the central values.

i. Where parameter values are proportions or percentages expected to be within the bounds of 0 to 1, the beta distribution is used in absence of alternative distribution information
ii. We suspect that the lower and upper bounds (0.09 and 0.13) that are presented in the table have been interpreted by the ERG as the parameters of the beta distribution, which is not the intended interpretation. The beta distribution parameters are available in the model.
iii. The majority of parameters that do not have uncertainty distributions (for example, A&E visit per week [BAT]) were used in calculations to produce other parameters (for example, Up to week 12 - A&E visit [for the relative adjustment between fedratinib resource use]) which were given uncertainty distributions. It was a concern that compounding uncertainty distributions would lead to erroneous or clinically implausible values being used in the sensitivity analysis. Other than those parameters, the compositions of BAT were fixed to maintain the proportions observed in the original sources while the user specified a desired proportion of ruxolitinib use.

We have re-examined the normal distributions; aside from deviations which could be attributed to rounding, we found no inconsistencies between the normal distributions confidence intervals and the mean.

A24. Appendix F.1.2, Table 46: Please comment on the observation that there is a dose related effect of survival with 500 mg fedratinib being similar to placebo.

It is acknowledged that the death rate in the fedratinib 500mgs arm was similar to the placebo arm. This could be due to similarity in baseline characteristics[14], noticeably greater percentage high risk status in fedratinib 500mgs ([redacted]) and placebo ([redacted]) arms compared to fedratinib 400mgs ([redacted]) arm but it could also be due to other factors. To provide an answer would be speculative.

The safety signals observed in patients receiving fedratinib 500 mg are not considered to be worse than placebo, although the fedratinib 400 mg dose has been submitted for regulatory filing given it represents an optimised risk/benefit profile.

Section B: Clarification on cost-effectiveness data

Stopping rule

B1. Priority. In CS (page 98), the company states that a 24-week treatment continuation rule for fedratinib can be implemented in the model. The stopping rule is not considered in the base-case, but presented as a scenario analysis.

i. Please confirm that no stopping rule will be usually applied in clinical practice, and that no such stopping rule is present in the expected licensing (compared with ruxolitinib)?
ii. In the scenario analysis, the stopping rule only affect costs, but not outcomes. Please clarify why the stopping rule is not expected to affect outcomes and whether this is in line with TA386.
iii. Please provide an analysis where outcomes are affected by the stopping rule, should the company consider the stopping rule to be relevant for fedratinib.

i. The fedratinib SPC states that [redacted]". As this isn't a definitive stopping rule, the stopping rule was not presented in the base-case, in line with NICE guidance. However, it does suggest that UK treatment guidance on discontinuation should be adhered to. It was also confirmed at the advisory board that a stopping rule would be used if patients had not responded at week 24.

ii. When the stopping rule is enabled in the model, incremental QALYs are lower, because patients transition to BAT sooner. Therefore, outcomes are impacted in the scenario analysis.

iii. Outcomes are already impacted by the stopping rule in the model.

Model structure and population

B2. Priority. A patient level simulation approach is used and it is stated in the CS (page 87-88) that the structure is similar to that used in TA386. Overall Survival (OS) in the model is modelled independently from other outcomes (TTD and response rate).

i. Please clarify what are the similarities and differences between TA386 and CS?

ii. The CS (page 88) further justify this approach to account for memory and transition between subsequent health states. Please clarify how this is accounted for in the model given that outcomes are modelled independently from each other.

iii. Please comment on the value of separating responder and non-responders in the model if no stopping rule is assumed and OS is modelled independently of response status?

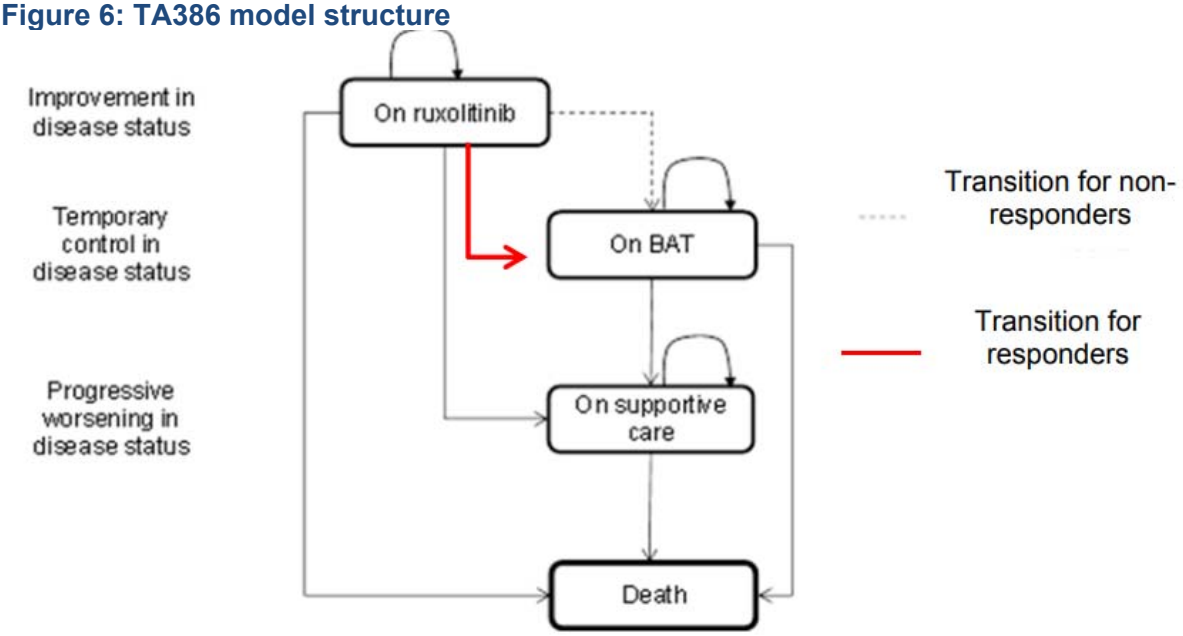
iv. The CS (page 96, 116) suggests that assuming the same duration of response (DoR) for both arms is conservative and acts as a waning of treatment effect. Please clarify how this acts as waning of treatment effect, when DoR is modelled independently from overall survival, and overall survival is taken from two separate sources (with parametric function fitted).

i. A comparison between the CS model and TA386 model is provided in the table below:

	TA386	CS model
Model type	Individual patients discrete event simulation with lifetime horizon	Same as TA386
Health states	On Ruxolitinib On BAT On supportive care Death (Leukaemic transformation counted as an adverse event)	Treatment states On JAKi (Fedratinib) On BAT Assessment states One-off response assessment event (included in both models but not described as a 'state' in TA386) Progressed states Acute myeloid leukaemia End of life states Palliative care Death
Model structure	See Figure 6 below. Ruxolitinib patients are categorised into 4 groups based on their outcomes at 24 weeks: Responders Non-responders Early discontinuation Early death The discontinuation of ruxolitinib after 24 weeks is dependent on the response assessment BAT patients discontinue according to modelled TTE, moving to supportive care and death	Fedratinib patients also have the 4 outcomes as listed in TA386, and discontinuation of fedratinib after 24 weeks is dependent on response assessment Supportive care is not a state, therefore patients on BAT remain until EOL states.
	A proportion of patients were modelled to die on treatment. For patients who discontinued the duration alive following discontinuation was modelled based on observed survival in the COMFORT-II trial using the difference in area under the curve between discontinuation and OS.	Discontinuation and OS are modelled independently. As such, death was used as a censor in the estimation of parametric models for TTD.
Time on treatment	No formal stopping rule was applied to patients receiving BAT Time to discontinuation was implemented by having patients transition to "supportive care" in the final 30% of time on BAT.	Once a patient received BAT, they remain on BAT until EOL.
	Ruxolitinib non-responders at 24 weeks were subjected to stopping rule – these patients moved to BAT state	Fedratinib non-responders discontinued at a rate calculated from JAKARTA-2 data

Response	Response is not modelled for patients receiving BAT	Patients receiving BAT are able to respond in the model
	HRQoL gains for patients with response are applied until discontinuation of ruxolitinib	Patients can lose response but remain on treatment in the model. HRQoL gains for patients are only experienced whilst the patient responds.

Figure 6: TA386 model structure



BAT, best available therapy.

i. The term “memory” in economic modelling typically refers to when the experience of a previous health state alters the patient experience in the current health state. This is applied in the current model following the response assessment health state, as response influences time on treatment and utilities. It is understandable that the ERG raises this query, given that the response assessment states are instantaneous health states.

   Furthermore, memory is not used in the model to link overall survival to other outcomes such as response and discontinuation. This was deemed appropriate given that the limited data available was not sufficient to produce a relationship between these outcomes, and that there would be considerable uncertainty over how this would translate to the BAT arm given that JAKARTA-2 is a single-arm trial. The current approach allows direct calculation of the overall survival from the JAKARTA-2 KM and the selected

BAT KM extrapolations, as opposed to indirect calculation via other events. This is especially important considering how OS parameters for both BAT and fedratinib are the most influential parameters according to sensitivity and scenario analyses.

ii. Separating out the responders and non-responders in the model allows a utility benefit to be applied for responders. Additionally, the application of time to treatment discontinuation calculations is dependent on response, so costs are influenced in this way. Finally, in addressing clarification item B6, the BAT composition following fedratinib has been split for responders and non-responders – to reflect the requested scenario that fedratinib patients may continue to receive fedratinib due to a lack of alternative treatment options. Please see the response to B6 for more detail.

iii. HRQoL benefits are applied to both fedratinib and BAT patients equally if a patient is modelled to respond. It would likely be inappropriate to assume that the HRQoL benefit should be applied until end of life. The JAKARTA-2 and JAKARTA studies reported that loss of response was often reported before treatment discontinuation, which has been interpreted here as treatment effect waning. As such, there is evidence that applying the HRQoL benefit whilst off-treatment may be inappropriate. Therefore, DoR is modelled as a separate outcome, and the HRQoL benefit is applied as long as patients have a response.

B3. Priority. A key input in the model (OS for BAT) is taken in people who discontinued ruxolitinib and are no longer treated with ruxolitinib. However, the company appear to consider a population that is relapsed/refractory where the majority of patients (89%) continue ruxolitinib treatment and therefore do not discontinue ruxolitinib.

i. Please clarify what is the population entering the economic model. Please clarify how inputs in the model match the population considered.

ii. The CS (page 98) states that “ruxolitinib alone is not considered a relevant comparator but is instead included as part of the basket of treatments in BAT. This attempts to ensure alignment between costs and efficacy inputs (See Section B.3.6.1).” Please clarify how costs and efficacy inputs are aligned?

iii. Please clarify the relevance of assessing response for ruxolitinib if the population of interest in the economic model is people who are maintained on

treatment and did not discontinue treatment because of suboptimal response and lack of effective therapies?

i. The population entering the model are patients with intermediate-2 or high-risk primary or secondary MF, who have been treated with ruxolitinib and are refractory/relapsed or are intolerant to ruxolitinib. The inputs and how they are relevant to the base case population are presented in the table below:

Input	**Relevance of source to population**
Baseline characteristics	Taken from JAKARTA-2, which contained population of interest
Response	Indirect treatment comparison between JAKARTA-2 ITT patients (containing 16/97 intermediate-1 risk patients) and SIMPLIFY-2 BAT patients. SIMPLIFY-2 study population was those with MF that had been treated with ruxolitinib. It was not possible to do this analysis for the intermediate-2/High-risk population only. It was assumed that the relevant difference between BAT and fedratinib response would be the consistent for the ITT population and the base case population.
Duration of treatment	Taken from JAKARTA-2 trial (ITT and int-2/high-risk populations both included as options), split by response post-24 weeks.
Duration of response	Taken from JAKARTA-2 trial (ITT and int-2/high-risk populations both included as options), used equally between treatment arms for responders.
Overall survival	Taken from JAKARTA-2 trial for fedratinib patients (ITT and int-2/high-risk populations both included as options). For BAT, the data is extrapolated from available literature sources because an ITC was not feasible.
Resource use	Sourced from HMRN and TA386. These sources are for a first-line population, however, the values were the most relevant source that was available.
BAT composition	Taken from SIMPLIFY-2 trial (in line with ITC) which had a similar patient population to JAKARTA-2.
Adverse events	Taken from JAKARTA-2 trial for fedratinib patients. BAT adverse event data was taken from available studies (SIMPLIFY-2 used in base case to align with ITC and BAT composition)

ii. As shown in the table above, the proportion of patients who respond in the BAT is informed by an ITC between JAKARTA-2 and SIMPLIFY-2 trials. The adverse event data and the composition of BAT was also taken from SIMPLIFY-2. Utilities are impacted by the proportion of patients who respond

and the frequency of adverse events. Costs are impacted by BAT composition and adverse events. By using the same source across these inputs, we attempt to align the costs and efficacy inputs for the BAT arm.

iii. It was relevant to assess the response for the BAT arm because the SIMPLIFY-2 BAT arm results showed that response for these patients did occur in small numbers (~6%) and it would have been inappropriate to exclude that treatment benefit. Assessing response for BAT in the model allows for a comparison of outcomes for patients receiving either fedratinib or BAT.

B4. Priority. In the economic model, overall survival (OS), time to treatment discontinuation (TTD) and duration of response (DoR) are sampled independently from each other. This leads to inconsistencies for TTD between the model predictions and the TTD parametric function used as shown in the Figure below (with transition to AML set to “No” [Sheet “Control”, Cell J218] and assumption of 0% receiving palliative care) generated by the ERG (AIC).

i. Please clarify and correct to ensure that the TTD predicted by the model matches the TTD curve used for both responder and non-responder. Please note that this inconsistency will also occur for DoR and need to be corrected.

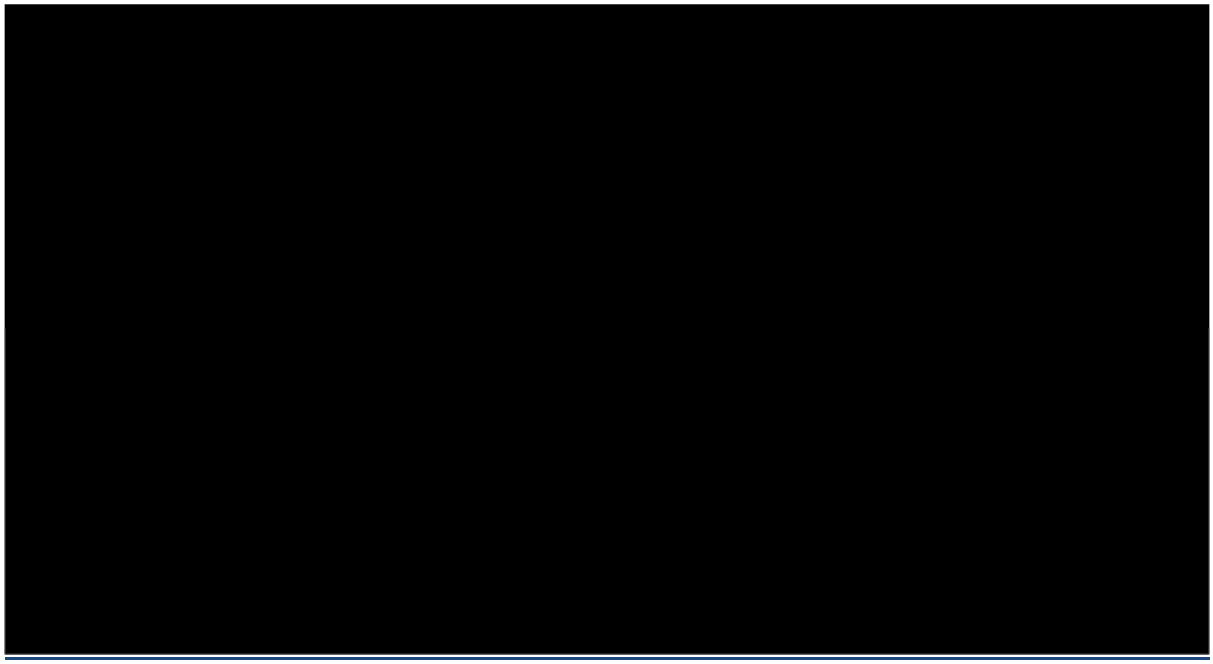


Figure produced by ERG (AIC)

i. In the knowledge that the economic model would estimate times-to-events independently, when parametric curves were fit to TTD and DoR data, death

was treated as a censor. Following this, it is an expected consequence that deaths in the economic model will lead to a lower time on treatment than the initial parametric curve.

If deaths were not censored in the original TTD/DoR curve estimation, then the impact of death on TTD/DoR would be double-counted in the economic model.

Therefore, the model output TTD and DoR curves are lower than their respective input parametric curves, which is what is observed in the graph generated by the ERG and is an intended mechanism for the model.

B5. Priority. When running the model with 10,000 patients, excluding transition to the AML health state [Sheet "Control", Cell J218], 0.44% (n=22/5037) of responders (Column K in "FED" sheet) and 0.37% (n=9/2410) of non-responders (Column L in "FED" sheet) discontinue treatment before Week 24. This is not consistent with the company's description of the economic model as people included in these two groups are on treatment and alive at Week 24. Please clarify if this is an error in the economic model and please amend the model accordingly.

Thank you for identifying this. There was an error in the economic model and it has been amended accordingly. The error did not impact the first 1000 patients in the model, therefore the base case results were not affected.

B6. Priority. The CS (page 17-18, 98, 155, 169) states that patients are continued on ruxolitinib despite achieving a suboptimal response in the absence of other targeted therapeutic options. In contrast, in the CS (page 132, Section B.3.3.4) patients entering the fedratinib arm are allowed to discontinue treatment (fedratinib) as per the trial discontinuation and move to the BAT arm (excluding ruxolitinib). This is inconsistent with the rationale provided in the CS (page 17-18, 98, 155, 169) that patients would remain on treatment as no other targeted therapeutic options are available.

i. Please clarify why patients on fedratinib with a suboptimal response would stop treatment, whilst patients on ruxolitinib remain on treatment for life.
ii. Please amend the model to reflect that patients initiating fedratinib would remain on treatment despite suboptimal response due to the absence of alternative targeted therapy (as per the assumption used for ruxolitinib). Patients switching from ruxolitinib to fedratinib have no other effective treatment options and therefore should continue to receive suboptimal fedratinib.

i. Data from clinical studies, which is supported by UK clinicians indicates that patients continue suboptimal ruxolitinib, therefore treatment was continued. In the absence of data to suggest that fedratinib would continue after patients have lost response, patients were modelled to move to BAT excluding a JAK inhibitor. This is also consistent to the approach taken in TA386.

   To allow a scenario to be consistent with the BAT arm, this has now been included in the model (see below).

ii. The model has been amended such that fedratinib can be included as a component of BAT. At the specified time-to-discontinuation for fedratinib, the revised model maintains the transition to BAT, which may now contain fedratinib.

   In addition, the BAT composition following fedratinib has been split in two: (1) a composition for patients who were initially responders and (2) a composition for patients who were not responders.

   If the assumption is made that patients can continue fedratinib beyond the current time-to-discontinuation, it may only be reasonable to assume this occurs in patients who initially responded. These are the patients who continue suboptimal ruxolitinib. It is expected that non-responders would discontinue fedratinib according to the time-to-discontinuation curves (or sooner with the stopping rule). The results of scenarios of continuing suboptimal fedratinib are presented below, along with the stopping rule.

Table 9: Incremental cost-effectiveness of scenarios where fedratinib is continued in responders only with the stopping rule applied.

Scenario	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER incremental (£/QALY)
Base case	8,545	0.848	0.615	13,905
25% of responders continuing fedratinib	8,884	0.848	0.613	14,505
50% of responders continuing fedratinib	18,584	0.848	0.613	30,341
75% of responders continuing fedratinib	28,284	0.848	0.613	46,178
100% of responders continuing fedratinib	37,984	0.848	0.613	62,014

Key: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Modelled overall survival

B7. Priority. The CS (Section B.3.3.3) report results from a systematic review of the literature to identify sources for OS post-ruxolitinib discontinuation. The company identified 13 studies, of which 4 are subsequently included in the economic model (CS, page 123).

i. Please clarify why overall survival data from the Mehra study (2016) were omitted from the economic model despite the KM being available in the poster (Figure 3 in Mehra et al, 2016) and the population in study similar to those in the other studies included. The CS (page 170) states that there is limited data for ruxolitinib versus non-ruxolitinib survival after ruxolitinib treatment failure. Please clarify why this study was not considered relevant.

ii. Please include an option in the model to use outcomes from the Mehra study (2016) separately for the subset of (a) patients receiving 2L ruxolitinib (Figure 3 in the Mehra paper – blue curve) and (b) patients receiving conventional therapy (Figure 3 in the Mehra paper – red curve).

iii. Please also include an option in the economic model to use data from the HMRN for OS post ruxolitinib discontinuation (CS, page 14, Figure 1)

i. Although the KM data was available for the Mehra study, the baseline characteristics for the patients who received ruxolitinib as a front-line therapy was not; primarily, the proportion of patients with intermediate-2 or high-risk classification was unknown. As such, the data is likely to contain patients that would not be eligible for fedratinib in the UK. Given that there were other studies, that were included as options within the model, which did report relevant baseline characteristics in some form, the Mehra study was not considered relevant as the outcomes could not be interpreted alongside any information on MF classification.

ii. The requested options have been added to the model, subject to limitations stated above.

iii. The HMRN OS data has been added as an option to estimate BAT survival in the economic model. The KM data shows a long tail owing to the low number at risk and the censoring at the later time points. Because of this, almost all of the parametric curves plateau at an unexpectedly high survival proportion. It is

clear from this that the presence of the long tail within the limited survival data has a negative impact on the appropriateness of this source.

B8. Priority. The CS (page 18 and 169) refer to data from SIMPLIFY-2 at 24 weeks to validate OS prediction for BAT (in people receiving continuous ruxolitinib) due to the limited evidence available, stating that 21% of patients died at 24 weeks in SIMPLIFY-2 and that this is consistent with the available BAT KM data.

i. Please clarify why the estimate from the KM (prior to cross over) at 24 weeks is not used (which show a different survival probability at Week 24).
ii. Please clarify how estimates from SIMPLIFY-2 at week 24 (KM before cross-over) for the BAT arm are consistent with the OS KM from Schain et al (2019) at Week 24 [used in the economic model].

i. The value used in the CS was the value reported in the reference material. The KM was not interpreted as we relied upon the authors to provide accurate interpretation of their own data.
ii. The KM is not consistent, the reported value of 21% patients dying at 24 weeks is consistent.

B9. Priority. The CS (page 169-170) states that OS for suboptimal treatment with ruxolitinib would be comparable to BAT OS and that the proportion of ruxolitinib in BAT would not significantly impact overall survival, and that this assumption was confirmed during the advisory board.

i. In the ad-board notes provided to the ERG, the clinical expert notes that people continuing ruxolitinib would have improved outcomes compared with conventional treatments despite being anaemic or on a suboptimal dose. Please provide a clear reference where this assumption was confirmed at the ad-board.
ii. In JAKARTA-2, patients with an expected life expectancy of less than 6 month were excluded from the trial. Studies considered for BAT (conducted at the point of discontinuation) have a sudden drop in survival as patients were not selected. This was recognised by the clinical experts during ad-board. Please clarify and comment on the implication of comparability between the JAKARTA-2 OS and OS from observational studies conducted at the point of discontinuation.
iii. Please clarify why OS at the point of ruxolitinib discontinuation is assumed to be the same compared with at the point at which a patient would be deemed relapsed/refractory but continued on treatment.

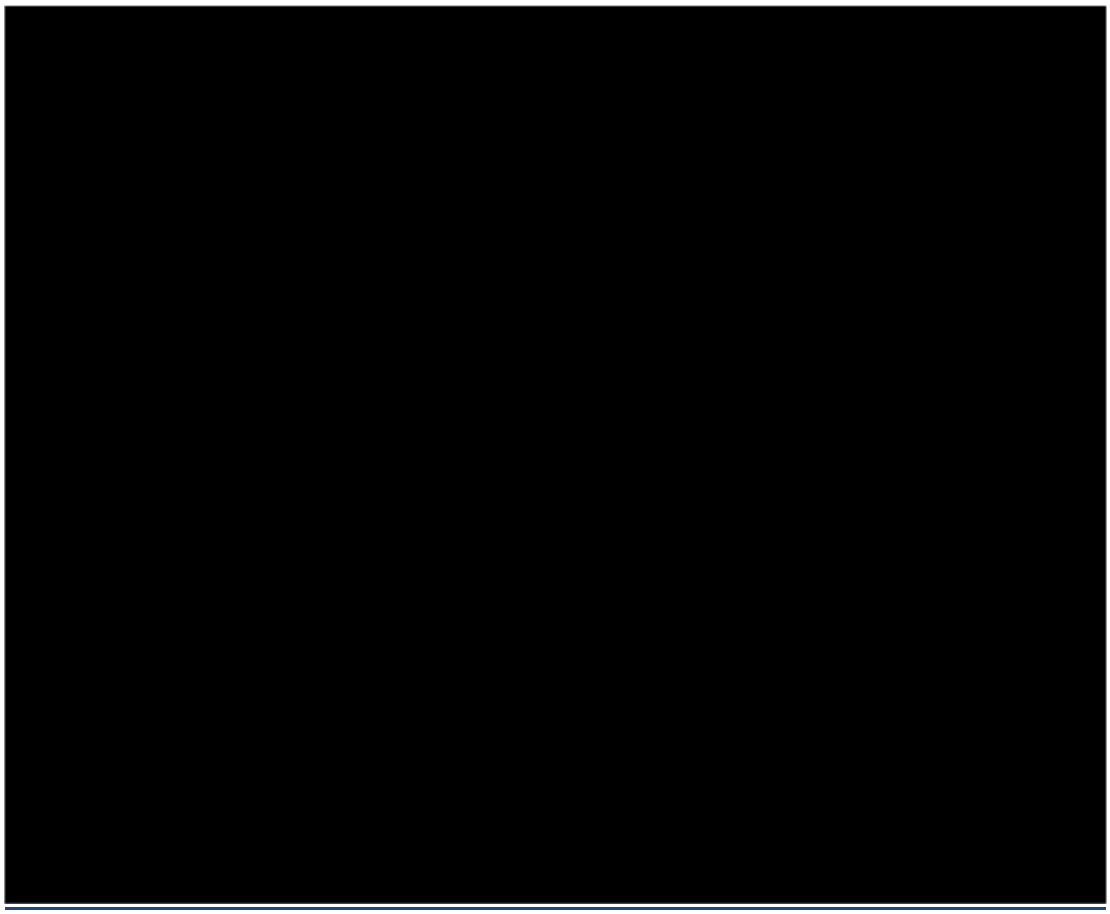
iv. Finally, in the CS (page 153) the company states that “Schain et al. presented results exclusively from Sweden and Norway, which clinicians decided may be inappropriate for a UK setting” when discussing the proportion of patients treated with ruxolitinib. Please clarify why this study was therefore selected for OS when this was deemed inappropriate by UK clinicians.

i. It was stated in the ad-board that a small benefit may be observed for OS, however, given the small proportions of response in the BAT population with 89% ruxolitinib (~6% SVR and TSS response) it was concluded that this benefit would not be clinically meaningful. Based on these comments, the Weibull curve was selected as the parametric model to extrapolate the OS data from Schain et al because the outcomes for this curve were slightly higher than the OS estimated by the clinical experts
ii. It is acknowledged that this may be a source of potential bias in the results. This exclusion criterion was a not an objective criterion and given the lack of information on the number excluded by this criterion it would have been difficult to adjust for the comparison to BAT observational data.
iii. This assumption was made owing to a lack of data to inform the survival of the individual populations
iv. It is acknowledged that this sentence was worded poorly. The composition of BAT from Schain et al. 2019 was decided by the clinicians to be inappropriate for calculating the costs in the model. However, on page 125 of the CS it is stated that “The group indicated that the population most representative of those expected to receive fedratinib in UK practice was that of Schain et al (2019)”. This was in part because the use of first-line ruxolitinib in Sweden and Norway MF patients is only approved for patients with a risk status of intermediate-2 or above. Both of the BAT OS KM data from Kuykendall et al. 2017 and Palandri et al. 2019 included patients classified as intermediate-1 or low risk. It was assumed that the risk classification of the patients included would have more influence over the OS outcomes than the BAT composition. As stated above (answer to question i), clinicians concluded that any benefit to OS from receiving ruxolitinib as part of BAT would not be significant; moreover, an optimistic parametric model was selected to extrapolate Schain et al 2019 data which can help address concerns over BAT composition.

B10. Please provide KM for OS from the JAKARTA-2 trial (Excel format) for:

i. PMF vs. other type of MF
ii. Relapsed vs. Refractory vs. intolerant

These KMs were not provided originally because of the lack of clinical rationale for separating out the populations and the limited data available. The graphs for the subgroups are provided below. There was overlap between intolerance and relapsed/refractory, so these are presented separately. There was no statistically significant difference between any of the populations and there was a large degree of uncertainty owing to the number of patients at risk.



Note: one of the patients had the reason for ruxolitinib failure stated as ‘other’, this patient was not included in this graph

Health-related quality of life

B11. Priority. Please provide a scenario analysis using the EQ-5D as done in TA386.

EQ-5D data is not available for the JAKARTA-2 trial. EQ-5D may not be appropriate for all patient groups or all populations. Limitations of generic measures in disease areas such as oncology are widely recognised; for example, psychometric analyses have indicated that the performance of EQ-5D in myelofibrosis (MF) is not ideal.[15, 16] Psychometric analyses of the performance of the EORTC QLQ-C30 against MF measures indicate that the EORTC QLQ-C30 captures functioning and some generic symptom problems. However, EORTC QLQ-C30 does not cover MF-specific symptoms (such as weight loss, itching, and night sweats) and is not as responsive

as the MFSAF over time. The myelofibrosis 8 dimension (MF-8D) was developed as a condition specific preference-based measure from the MFSAF version 2.0 and the EORTC QLQ-C30 that captures the HRQOL of patients with MF and overcomes some of the concerns related to using the EQ-5D and EORTC QLQ-C30.

B12. Priority. Please clarify why a mixed effect model was used for utility values, compared with an alternative model that better represents the data. Please also clarify why Gender is included in the regression model to estimate utility values (CS. Table 55) and whether there is evidence that utility is predicted by Gender

i. Please include an option in the model to use utility values when Gender is removed.

No specific utility analyses NICE TSD guidance exists. Mixed effects models have been fitted for utility values given they are repeated measures data. Alternative models, such as those presented in Alava et al. (2012),[17] are developed primarily to address three key issues in the utility values: floor effects, ceiling effects and multimodal distributions. Whether these models are practically beneficial for the type of utility values in JAKARTA-2 is unclear. Utility values from JAKARTA-2 do not display a multimodal distribution nor is there a mass of observations at 1, histograms of the utility values can be found in Figure 7. Residual diagnostics of the mixed effect models suggest that the residual assumptions of the mixed effect models are reasonable. Consequently, the mixed effect model is used for utility values.

Figure 7: Histograms of MF-8D (left) and EORTC-8D (right) utility values from JAKARTA-2

Gender was included as a covariate in the utility mixed effect models. There was evidence in the data to suggest gender had a small effect on utility in exploratory analyses and univariate and multivariate regression models, see Table 10. Alternative mixed effect models have been fitted for the utility values with gender removed from the covariates, see Table 11 and Table 12. An option has been added to the economic model to use the utility values models without gender as a covariate in the mixed effect model.

Table 10: Mixed effects model of MF-8D and EORTC-8D utility values using baseline and gender

Parameters	Int2/high risk population			ITT population		
	Estimate	SE	p-value	Estimate	SE	p-value
MF-8D						
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline MF-8D	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gender - Female	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gender - Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EORTC-8D						
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline EORTC-8D	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sex - Female	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sex - Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Note: Ref, reference group, SE, standard error.						

Table 11: Parsimonious mixed effects model – Int2/High risk population

Parameters	Spleen response model			Symptom response model			Spleen and/or symptom response		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
MF-8D									
Intercept	[redacted]								
Baseline MF-8D	[redacted]								
Spleen non-response	[redacted]								
Spleen response	[redacted]								
Symptom non-response	[redacted]								
Symptom response	[redacted]								
Spleen and/or symptom non-response	[redacted]								
Spleen and/or symptom response	[redacted]								
EORTC-8D									
Intercept	[redacted]								
Baseline EORTC-8D	[redacted]								
Spleen non-response	[redacted]								
Spleen response	[redacted]								
Symptom non-response	[redacted]								
Symptom response	[redacted]								
Spleen and/or symptom non-response	[redacted]								
Spleen and/or symptom response	[redacted]								

Note: Ref, reference group, SE, standard error.

Table 12: Parsimonious mixed effects model – ITT population

Parameters	Spleen response model			Symptom response model			Spleen and/or symptom response		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
MF-8D									
Intercept	[redacted]								
Baseline MF-8D	[redacted]								
Spleen non-response	[redacted]								
Spleen response	[redacted]								
Symptom non-response	[redacted]								
Symptom response	[redacted]								
Spleen and/or symptom non-response	[redacted]								
Spleen and/or symptom response	[redacted]								
EORTC-8D									
Intercept	[redacted]								
Baseline EORTC-8D	[redacted]								
Spleen non-response	[redacted]								
Spleen response	[redacted]								
Symptom non-response	[redacted]								

Parameters	Spleen response model			Symptom response model			Spleen and/or symptom response		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
Symptom response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen and/or symptom non-response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Spleen and/or symptom response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Note: Ref, reference group, SE, standard error.									

Figure 8: MF-8D model diagnostic plots – ITT population – mixed effect model with covariates for baseline utility and spleen and/or symptom response

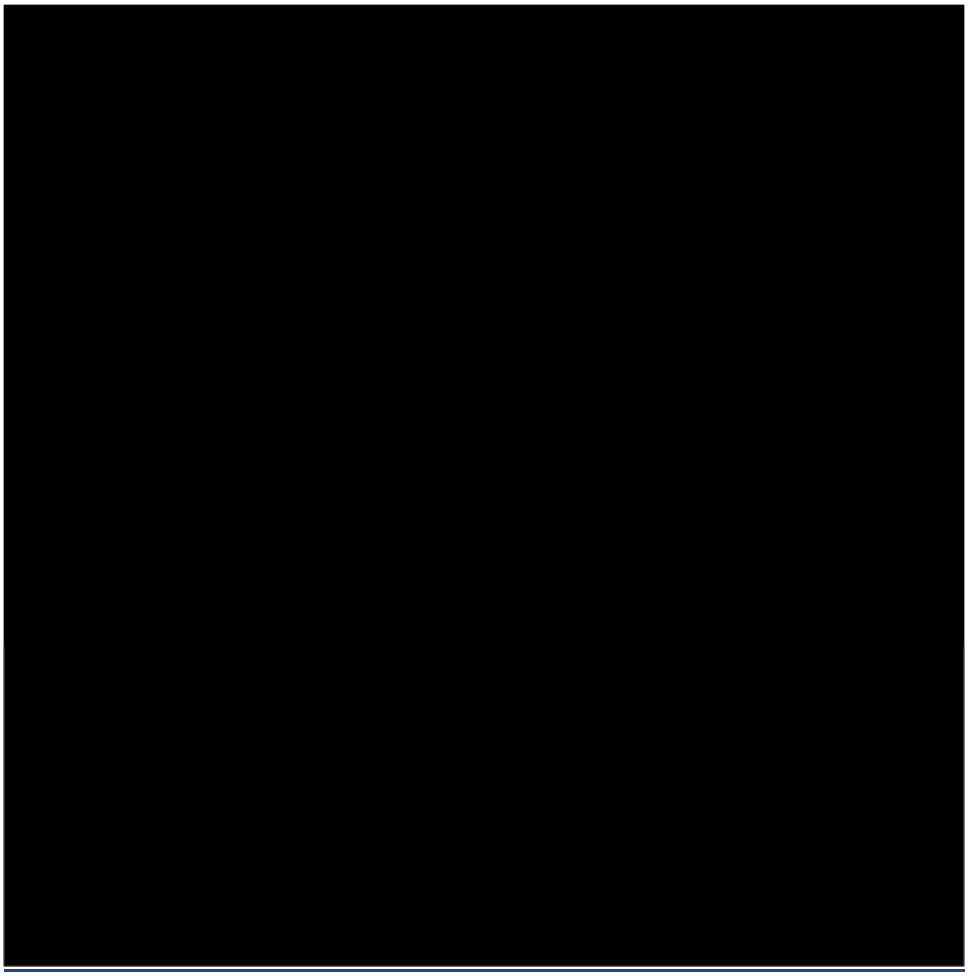


Figure 9: EORTC-8D model diagnostic plots – ITT population – mixed effect model with covariates for baseline utility and spleen and/or symptom response

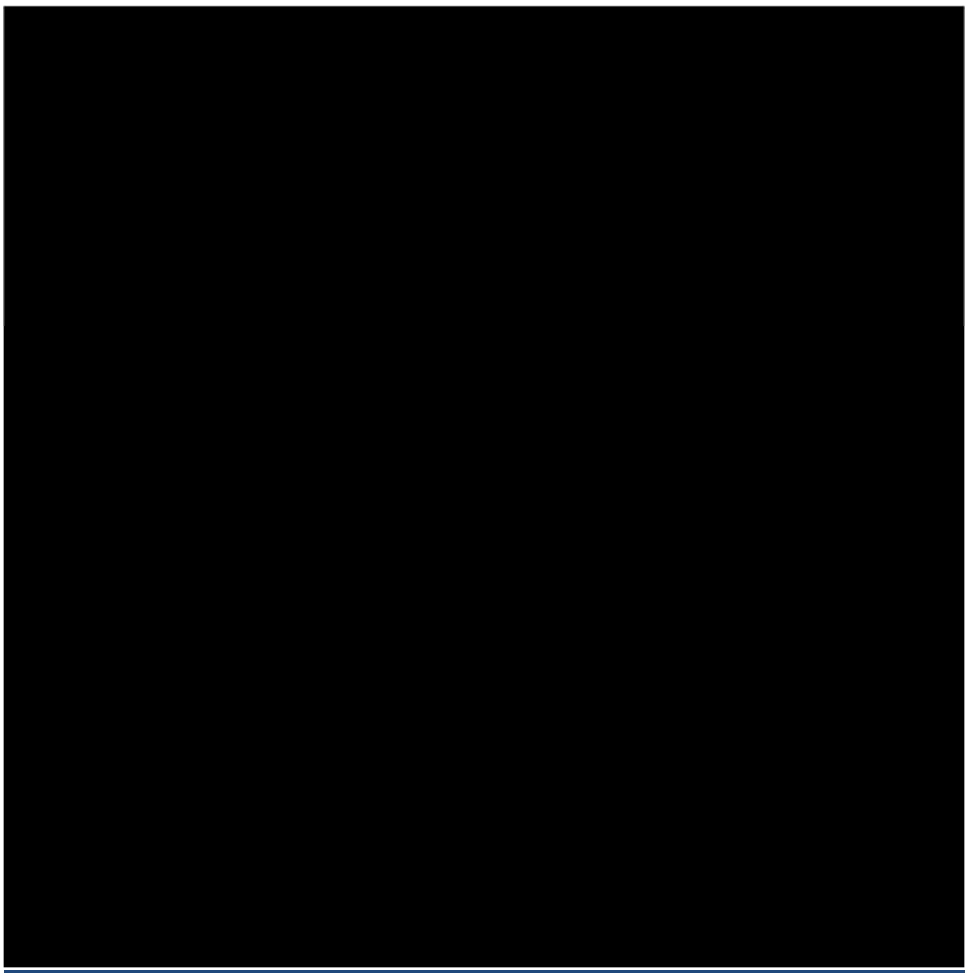


Figure 10: Figure 11: MF-8D model diagnostic plots – Int2/High risk population – mixed effect model with covariates for baseline utility and spleen and/or symptom response

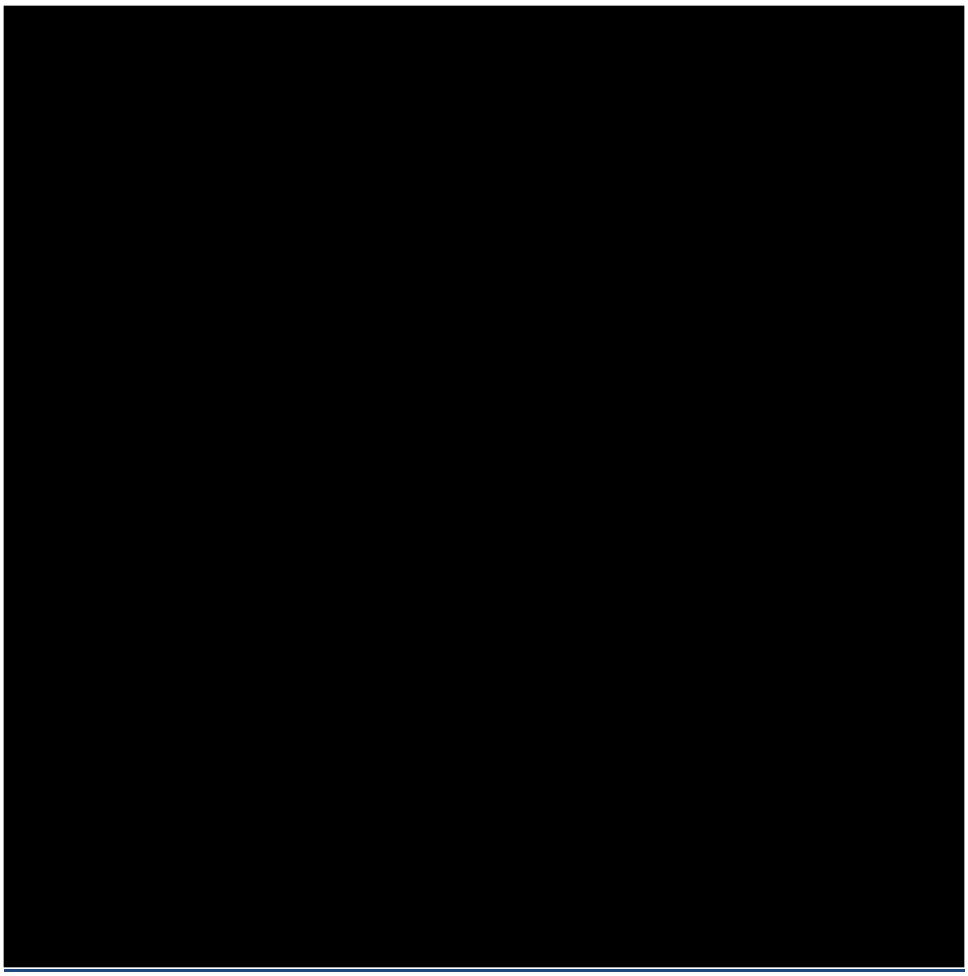
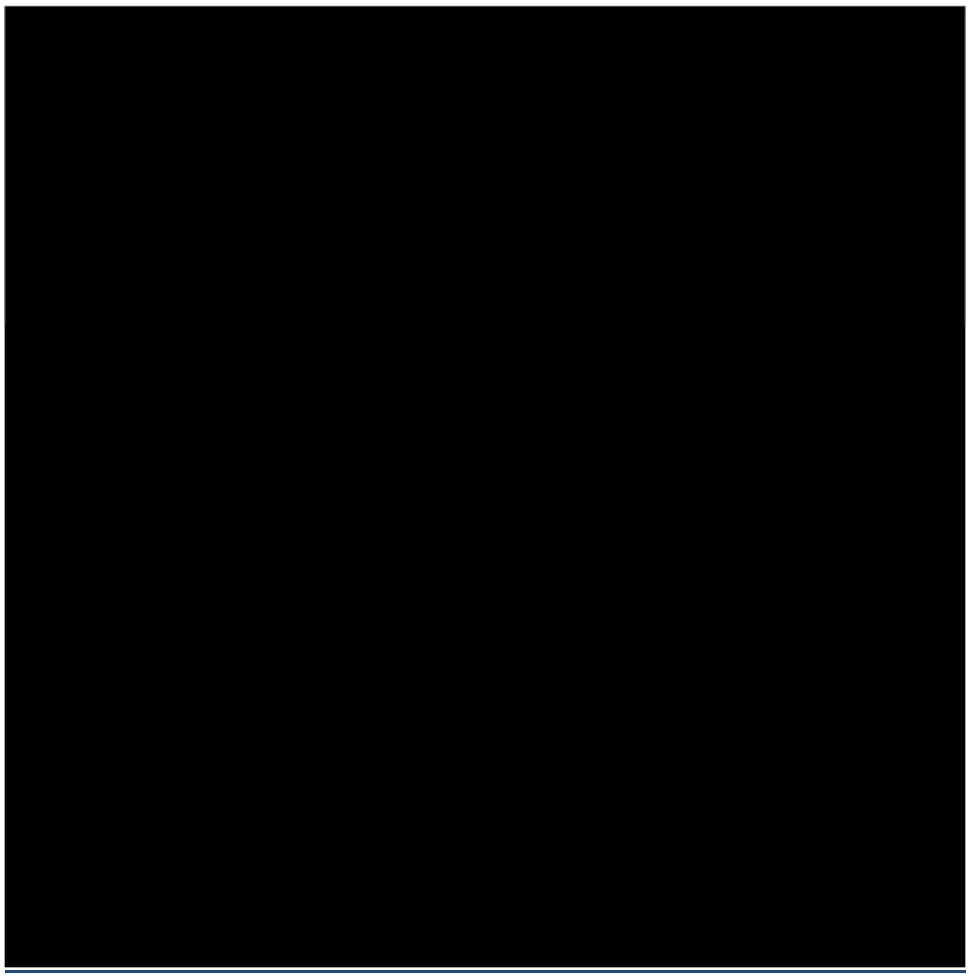


Figure 12: EORTC-8D model diagnostic plots – Int2/High risk population – mixed effect model with covariates for baseline utility and spleen and/or symptom response



Response rates and TTD

B13. Priority. The CS (section B.3.3.2.2) states that data on response rate in the economic model for BAT are taken from the SIMPLIFY-2 and PERSIST trials. Please clarify how the population included in SIMPLIFY-2 trial are comparable to the population included in the JAKARTA-2 trial. In particular;

i. Patients in the SIMPLIFY-2 trial appear to be intolerant only and had to either require red blood cell transfusions while on ruxolitinib or ruxolitinib dose reduction with at least one of grade 3 thrombocytopenia, anaemia, or bleeding. Please clarify how this population fits with the resistant population included in the JAKARTA-2 trial.
ii. No washout period was included in the SIMPLIFY-2 compared with the JAKARTA-2 trial. Please clarify how this would affect response rate.

i. There is no internationally recognised criteria for intolerant or resistance, which means that assignment to these patient groups is open to interpretation and there could be overlap. Clinical advice indicates that there can sometimes be difficulty in defining intolerance, such that patients can have a mixture of both intolerance and resistance.
ii. The JAKARTA-2 protocol mandated a short washout period, while the SIMPLIFY-2 protocol did not. We cannot speculate on the response rates observed or derived from the washout period. However, A short washout period was observed in PERSIST-2, the results for which are suggestive of washout not having a marked effect on response. However, this must be interpreted with caution due to the different study population. To note, the FDA have mandated washout periods for all future myelofibrosis studies.

B14. The CS (page 132) states that the time to discontinuation in the group of people with early discontinuation is estimated using a uniform distribution between 0 and 24 weeks.

i. Please clarify why direct data from the trial are not used?
ii. Please provide the KM (in Excel) for those patients

i. The direct data were not used because the clinical hold had impacted discontinuations for the base case population, leading to uncertainty and small numbers at risk. It was felt that the model should be flexible enough to consider different potential proportions of early discontinuations, which is what the current model approach allows.
ii. The KMs have been included in the Excel model and are displayed below. The uniform distribution applied in the model is supported by the linearity of the KM data.

B15. Please clarify whether the duration of response (DoR) [CS, page 117] is calculated from the time to response or from week 24 onward or another time point. Please clarify how this is applied in the economic model and please confirm this has been applied appropriately (from time to response or week 24 as appropriate)

In the estimation of parametric curves for DoR, times were calculated from the time of response. Only patients who responded at week 24 were included in this estimation, to ensure alignment with the economic model which only assigns response to week 24 responders.

However, in the economic model, the DoR curves are applied from week 24 (not from the time of response). This represents an inconsistency in the economic model as pointed out by the ERG.

Therefore, in a revised model, an option is included which uses DoR data that has been re-based to commence from 24 weeks. The impact on the data is shown in Figure 13. There is a visible shift in the KM when re-based to 24 weeks, because many responders experienced response prior to week 24.

Figure 13: Comparison of duration of response Kaplan-Meier plots (original analysis vs. re-based analysis)

Of note, 3 responders (in both the ITT and int2/HR population) were censored before 24 weeks, and so were re-based with negative values (the maximum difference was 12 days). Because these observations were censors, not events, there was no impact on the KM plot, but the values were changed to 0.1 for the parametric fits as negative values would not be compatible.

Clinical expert opinion

B16. Priority. The CS provides notes from the UK advisory-board (ad-board) in the reference list pack to support some of its arguments. Please provide the slides presented during this ad-board, as well as the full report.

In response, we have shared the advisory board slides and the supportive tool used to validate OS extrapolations.

B17. Priority. The CS (page 97, 150) states that clinical feedback indicated that BAT and 'supportive care' were equivalent.

i. Please provide a clear reference supporting this statement.

The intended interpretation here is that patients who are on BAT in the setting of relapsed/refractory or intolerant to ruxolitinib are on a therapy that can be considered supportive care, in that it does not achieve a trial endpoint.

Mesa 2014 reports that:[18] *'patients who received BAT in COMFORT-II appeared to fare no better than patients who received placebo in COMFORT-I, and these findings illustrate that conventional therapeutic alternatives for patients with myelofibrosis do not alleviate the symptom burden of the disease in a meaningful way, underscoring the need for better treatments.'*

B18. The CS (page 124) states that "prior to the UK advisory board held for fedratinib, clinician attendees (N=7) were asked to consider and provide their expectations of survival in the post ruxolitinib population for those treated with BAT and those treated with fedratinib….The averages of these estimates are shown in Table 48".

i. Please provide a copy of the questionnaire sent as well as all the individual responses (anonymised).

In response, we have shared a copy of the questionnaire and the anonymised individual responses.

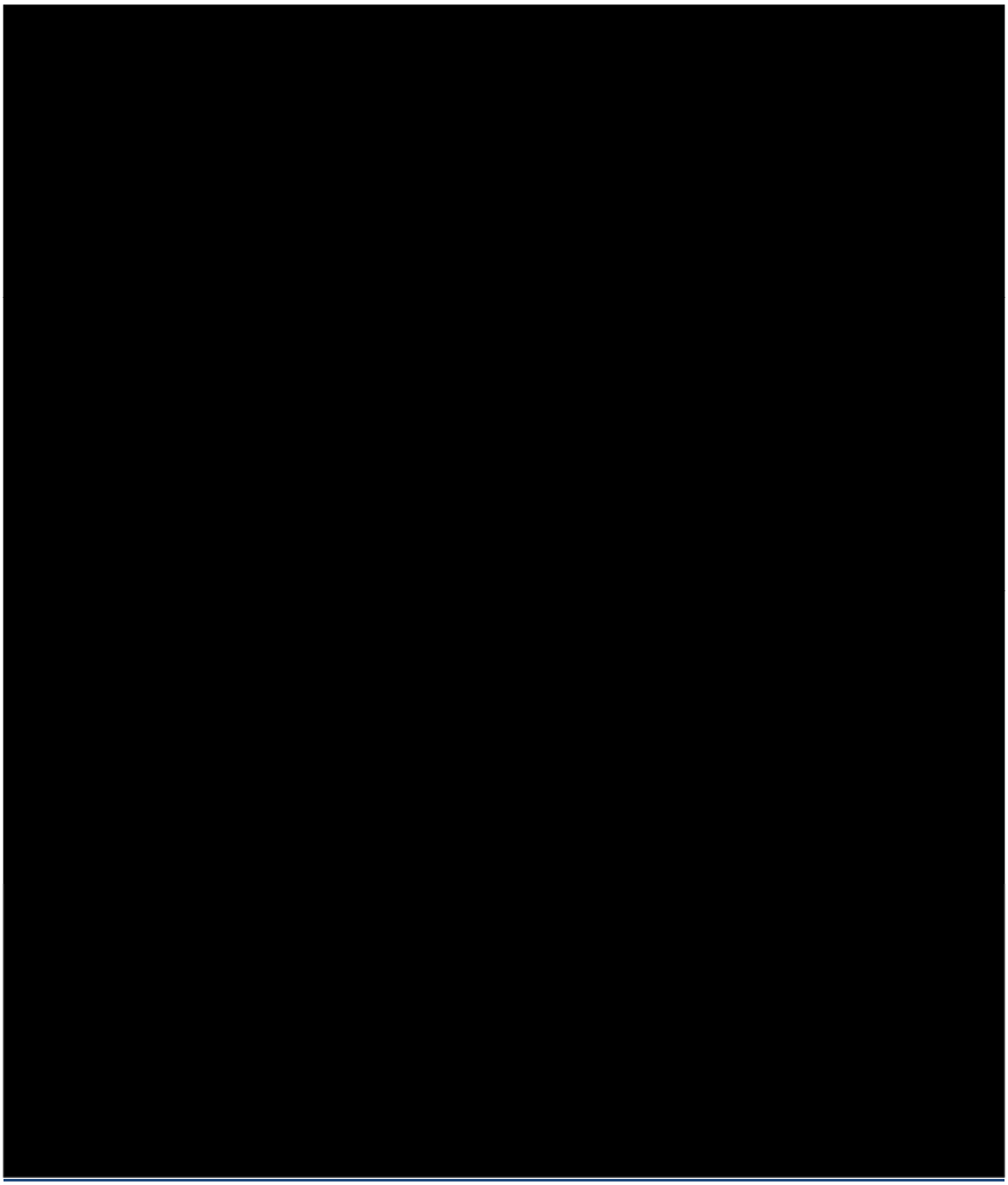
B19. The CS (page 128) states that the Gompertz distribution is used to represent OS for the Int2/high risk population following clinical opinion. In the ad-board notes, it is suggested that the generalised gamma is more clinically reasonable.

i. Please clarity this inconsistency why the company used the Gompertz when clinical experts considered the generalised gamma to be more clinically appropriate.

The final report from the advisory board shows the consensus values from the clinicians for the OS estimates of fedratinib. Within this advisory board the clinicians suggested that the intermediate-2/high-risk population should be the focus because it was agreed that this was representative of how fedratinib will be used in clinical practise. However, only ITT data were shown to the clinicians at the time, as opposed to intermediate-2/high-risk data.

The Gompertz and generalised gamma curves were most similar to the consensus values (see Figure 14). After the advisory board, curves were fit to the intermediate-2/high-risk data, and the new Gompertz curve was most similar to the ITT generalised gamma and Gompertz curves.

Figure 14: ITT advisory board consensus for fedratinib OS



Other data sources

B20. In Appendix L6 the company states that “Palandri et al. (2019) provides supportive evidence of prolonged survival post-ruxolitinib with ‘novel agents’ such as fedratinib when compared to ‘conventional agents’ (Figure 13). The remaining sources of OS did not report survival for novel agents. “. Please comment why this source is deemed inappropriate for BAT (consisting of mostly ruxolitinib) when about 35% in this study (n=11/31) received ruxolitinib compared with 3% (n=1/31) receiving fedratinib.

The patients who started ruxolitinib consisted of 52.4% intermediate-1 risk patients. Although the risk classification of the patients who were treated after ruxolitinib was not stated, given the starting population, it is highly likely that the proportion of intermediate-1 patients in this population would have been inappropriate to compare to the model base case population. In addition, the population in Palandri et al. excluded patients who were in blast phase, whereas the JAKARTA-2 study did not include blast phase as part of the exclusion criteria. Given the Palandri et al shows that patients in blast phase have poor prognosis, it is likely that the exclusion of these patients would make them incomparable to the model patient population. The unknown nature and efficacy of the investigational therapies administered to the majority of patients also contributed to the decision that this source was inappropriate.

Adverse events

B21. The CS (page 143, Section B.3.4.4) states that the frequency of adverse events for patients treated with fedratinib is taken from the Int2/high risk subgroup only (n=81).

i. Please clarify why the frequency of AE is not calculated from the ITT population
ii. Please provide values (Table 59 & Table 60; page 144-145) for the ITT population, in addition to the mean weeks of exposure for the ITT population
iii. Please provide similar tables (frequencies of Adverse events and mean weeks of exposure) from the JAKARTA trial (conducted in 1L) for both the fedratinib and placebo arm.

i. The int-2/high-risk population was the base case population, so it was used for AE frequency as opposed to the ITT.

ii. Tables are presented below. The second table is provided in the CS model on the ‘Adverse Events’ sheet

Adverse event	Fedratinib AEs (JAKARTA-2 ITT [N=97])
Anaemia	[redacted]
Thrombocytopenia	[redacted]
Leukopenia	[redacted]
Splenomegaly	[redacted]
Cytopenia	[redacted]
Febrile Neutropenia	[redacted]
Leukocytosis	[redacted]
Neutropenia	[redacted]
Thrombotic thrombocytopenic purpura	[redacted]
Key: AE, adverse events; N, total patients.	

Adverse event	n	N	Source
Abdominal pain	[redacted]	[redacted]	JAKARTA-2 CSR
Arthralgia	[redacted]	[redacted]	JAKARTA-2 CSR
Asthenia	[redacted]	[redacted]	JAKARTA-2 CSR
Back pain	[redacted]	[redacted]	JAKARTA-2 CSR
Bronchitis	[redacted]	[redacted]	JAKARTA-2 CSR
Cough	[redacted]	[redacted]	JAKARTA-2 CSR
Diarrhoea	[redacted]	[redacted]	JAKARTA-2 CSR
Dyspnoea	[redacted]	[redacted]	JAKARTA-2 CSR
Fatigue	[redacted]	[redacted]	JAKARTA-2 CSR
Headache	[redacted]	[redacted]	JAKARTA-2 CSR
Nausea	[redacted]	[redacted]	JAKARTA-2 CSR
Oedema peripheral	[redacted]	[redacted]	JAKARTA-2 CSR
Pain in extremity	[redacted]	[redacted]	JAKARTA-2 CSR
Pyrexia	[redacted]	[redacted]	JAKARTA-2 CSR
Weight increased	[redacted]	[redacted]	JAKARTA-2 CSR

Key: CSR, clinical study report; n, number of patients with event; N, total number of patients.
Notes: Mean exposure to fedratinib in JAKARTA-2 was 0.539 years. Only adverse events with severity Grade ≥ 3 were considered.

iii. JAKARTA adverse events are presented below. The mean and median exposure by treatment arm for the entire treatment duration is also presented.

Adverse event	Placebo grade 3-4 AEs [N=95]	Fedratinib 400mg grade 3-4 AEs [N=96]	Fedratinib 500mg grade 3-4 AEs [N=97]
Haematological Adverse Events			
Anaemia	[redacted]	[redacted]	[redacted]
Thrombocytopenia	[redacted]	[redacted]	[redacted]
Leukopenia	[redacted]	[redacted]	[redacted]
Splenomegaly	[redacted]	[redacted]	[redacted]
Cytopenia	[redacted]	[redacted]	[redacted]
Febrile Neutropenia	[redacted]	[redacted]	[redacted]
Leukocytosis	[redacted]	[redacted]	[redacted]
Neutropenia	[redacted]	[redacted]	[redacted]
Thrombotic thrombocytopenic purpura	[redacted]	[redacted]	[redacted]
Disseminated Intravascular Coagulation	[redacted]	[redacted]	[redacted]
Other adverse events			
Abdominal pain	[redacted]	[redacted]	[redacted]
Arthralgia	[redacted]	[redacted]	[redacted]
Asthenia	[redacted]	[redacted]	[redacted]
Back pain	[redacted]	[redacted]	[redacted]
Bronchitis	[redacted]	[redacted]	[redacted]
Cough	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]
Dyspnoea	[redacted]	[redacted]	[redacted]

Fatigue	[redacted]	[redacted]	[redacted]
Headache	[redacted]	[redacted]	[redacted]
Nausea	[redacted]	[redacted]	[redacted]
Oedema peripheral	[redacted]	[redacted]	[redacted]
Pain in extremity	[redacted]	[redacted]	[redacted]
Pyrexia	[redacted]	[redacted]	[redacted]
Weight increased	[redacted]	[redacted]	[redacted]
Key: AE, adverse events; N, total patients.			

		Fedratinib	
	Placebo (N = 95)	**400 mg (N = 96)**	**500 mg (N = 96)**
Duration of Exposure (weeks)			
Mean (SD)	[redacted]	[redacted]	[redacted]
Median	[redacted]	[redacted]	[redacted]

Resource use and costs

B22. The CS (page 165) assumes that patients requiring thiamine testing and supplementation (200 mg daily) require a 90-day course at treatment initiation (baseline) and at treatment cessation (at the point of discontinuation).

i. Please clarify why patients are not treated continuously until treatment discontinuation and why supplementation is only given at the start and end of treatment.

ii. The CS (page 165) states that Thiamine dose may vary between 50mg – 300mg per day and assumes that patients requiring thiamine supplementation receive 200 mg daily. In the BNF, the recommended dose for adult is 200-300 mg daily. Please clarify.

i. An expected prescribing practice for thiamine supplementation alongside fedratinib was lacking in the literature, and so this was implemented as a simplifying assumption.

In the revised model we submit alongside these responses, the option is included to treat continuously until treatment discontinuation.

ii. The wider range was taken from the SPC for thiamine which includes treatment for patients with a mild deficiency (50 mg to 100 mg).

B23. Please include the impact of Wernicke's encephalopathy (WE) as an adverse event in the economic model.

In November 2013, a clinical hold was placed on the fedratinib program following the emergence of a potential signal of WE in fedratinib-treated patients. Fedratinib safety was then evaluated in 608 patients who received more than one fedratinib dose, including 459 patients with MF. Eight potential cases (1.3%) of Wernicke's encephalopathy were identified, and one case (0.16%) was fatal. Only one case was confirmed. All suspected cases were in the 500 mg fedratinib arm of JAKARTA. All 8 potential Wernicke's encephalopathy cases were associated with pre-existing malnutrition and weight loss and/or significant nausea and vomiting that were not adequately controlled.[19]

From fedratinib's first approval by the United States (US) Food and Drug Administration (FDA) on 16 Aug 2019 until 31 May 2020, an estimated [redacted] patients have been exposed to commercial fedratinib. [redacted] was reported in an 82-year-old female with a medical history [redacted] prior to the start of fedratinib treatment. Magnetic resonance imaging (MRI) was [redacted] and neurology was consulted and did not believe the patient had Wernicke's.

WE is not an expected adverse event for patients receiving 400 mg fedratinib. This in line with the JAKARTA-2 data and supported by the fact that thiamine levels are to be monitored for all patients considered for fedratinib prior to and during treatment (as per US PI/ draft SPC). Therefore, WE is not considered a relevant AE for this economic model.

Subgroup and definitions

B24. Please clarify why no subgroup analysis is conducted separately for patients that that are (a) relapsed, (b) refractory and (c) intolerant?

Efficacy findings in subpopulations relapsed/refractory versus intolerant are publicly available and provided in Appendix E.[8] Subgroup analyses in relapsed versus refractory patients were not conducted as there is no clinical justification to support separating out these populations. Additionally, any observations in these subgroups would be limited by the sample size (i.e. 47 refractory patients and 18 relapsed patients).

B25. In the JAKARTA-2 trial, the definition for resistance (relapse or refractory) is based on spleen volume only. Please comment on how this relates to clinical practice where symptom response is also a relevant measure for clinical benefit?

It is acknowledged that spleen volume and symptom response both play a significant role in determining clinical benefit in clinical practice. The JAKARTA-2 study was a single arm, Phase 2, non-randomised study which needed an objective endpoint. Spleen volume (measured by MRI/CT scan) enabled this objectivity. Symptom response, although relevant, can be more subjective.

Section C: Textual clarification and additional points

Identification, selection and synthesis of clinical evidence

C1. Appendices D.1.2. Study Selection, first paragraph. How many reviewers were involved in screening records at the title stage? How many reviewers were involved in screening records at the abstract stage? What proportion of all records were screened by two reviewers?

Studies were assessed for eligibility by two independent reviewers, with disagreements adjudicated by a third reviewer. This was applied at both title/abstract and full-text screening stage to ensure everything is quality checked

C2. Appendices D.1.3.1. Complete reference lists for included studies and excluded studies, first paragraph. Please provide a table of the studies excluded at the full-text

stage (n=326) with the reason for inclusion for each of the studies, along with a reference list for these studies

This is provided in the following document:

C3. Appendices D.1.3.1. Data extraction and quality assessment. First paragraph, how many of the items for data extraction were checked by the second reviewer and how was this undertaken? Second paragraph, please provide the reference for the Downs and Black QA instrument used

All the extracted data was quality checked independently by a second reviewer.
The reference for the Downs and Black QA instrument used is as follows:
Downs SH and Black N. (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 52(6): 377-384.

C4. Appendices D.1.4.3. Risk of bias of studies included in indirect or mixed treatment comparisons. First paragraph, Please provide the name of the quality assessment instrument used and the citation for this.

The NICE checklist was used to assess the quality of the included RCT studies (see Appendix D.1.3.2), this is referenced to:
National Health and Care Excellence (NICE). (2015) (Updated: April 2017) Single Technology Appraisal (STA): User guide for company evidence submission template. Available at: https://www.nice.org.uk/article/pmg24/chapter/4-Clinical-effectiveness#quality-assessment-of-the-relevant-randomised-controlled-trials. Accessed: July 2020.

C5. CS page 14 “The HMRN figure may not be considered representative of UK clinical practice, given that the estimated uptake of ruxolitinib following its approval and licencing was not observed in HMRN data (see Appendix N).” Please provide further supported information on this statement, including sources.

Given that ruxolitinib represents the only targeted therapy available for patients with myelofibrosis, it is highly plausible to expect an increase in its uptake following licencing for use in UK patients.[11] This trend was not observed in the HMRN data, as described in Appendix N:

Table 12: Start and end year of patients who were initiated ruxolitinib

	Ruxolitinib Start Year	Ruxolitinib End Year
Total	[redacted]	[redacted]
2012	[redacted]	[redacted]
2013	[redacted]	[redacted]
2014	[redacted]	[redacted]
2015	[redacted]	[redacted]
2016	[redacted]	[redacted]
2017	[redacted]	[redacted]
2018	[redacted]	[redacted]
2019	[redacted]	[redacted]

Key: HMRN, Haematological Malignancy Research Network.
Source: HMRN report[20]

Of [redacted] patients receiving ruxolitinib in the HMRN analysis, [redacted]% started ruxolitinib therapy in 2016. The proportion of patients starting ruxolitinib each year [redacted] in subsequent years. Conversely, ruxolitinib had received a positive recommendation from NICE in 2016, and whilst market share data for ruxolitinib is not publicly available, the lack of alternative therapies supports the assumption that ruxolitinib uptake increased during this period. In this context, given the small number of myelofibrosis treatment centres within the HMRN catchment area, the HMRN data may not be considered representative of UK clinical practice.

References

1. Tang D, Taneja A, Rajora P, et al. Overall Survival in Patients with Myelofibrosis Who Have Discontinued Ruxolitinib: A Literature Review. American Society of Hematology Washington, DC, 2019.
2. Mesa RA, Gotlib J, Gupta V, et al. Effect of ruxolitinib therapy on myelofibrosis-related symptoms and other patient-reported outcomes in COMFORT-

I: a randomized, double-blind, placebo-controlled trial. *Journal of Clinical Oncology*. 2013; 31(10):1285.
3. Harrison C, Kiladjian JJ, Al-Ali HK, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N Engl J Med*. 2012; 366(9):787-98.
4. Harrison CN, Schaap N, Vannucchi AM, et al. Janus kinase-2 inhibitor fedratinib in patients with myelofibrosis previously treated with ruxolitinib (JAKARTA-2): a single-arm, open-label, non-randomised, phase 2, multicentre study. *Lancet Haematol*. 2017; 4(7):e317-e24.
5. Verstovsek S, Kantarjian HM, Estrov Z, et al. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. *Blood*. 2012; 120(6):1202-9.
6. Celgene. ARD12181 (JAKARTA2): A Phase II, multicenter, open-label, single-arm study of SAR302503 in subjects previously treated with ruxolitinib and with a current diagnosis of intermediate or high-risk primary myelofibrosis, post-polycythemia vera myelofibrosis, or post-essential thrombocythemia myelofibrosis (clinical study report). 18 Dec 2018 2018.
7. Hussein K, Huang J, Lasho T, et al. Karyotype complements the International Prognostic Scoring System for primary myelofibrosis. *Eur J Haematol*. 2009; 82(4):255-9.
8. Harrison CN, Schaap N, Vannucchi AM, et al. Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the JAKARTA2 study using stringent criteria for ruxolitinib failure. *American Journal of Hematology*. 2020.
9. Celgene. Myelofibrosis Advisory Board Meeting Report. April 8th 2020 2020.
10. Schain F, Vago E, Song C, et al. Survival outcomes in myelofibrosis patients treated with ruxolitinib: A population-based cohort study in Sweden and Norway. *Eur J Haematol*. 2019; 103(6):614-9.
11. National Institute of Care Excellence (NICE). Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. Technology appraisal guidance [TA386]. 2016. (Updated: 23 March 2016) Available at: https://www.nice.org.uk/guidance/TA386/chapter/1-Recommendations. Accessed: 13 November 2019.
12. Kuykendall AT, Shah S, Talati C, et al. Between a rux and a hard place: evaluating salvage treatment and outcomes in myelofibrosis after ruxolitinib discontinuation. *Ann Hematol*. 2018; 97(3):435-41.
13. Palandri F, Breccia M, Bonifacio M, et al. Life after ruxolitinib: Reasons for discontinuation, impact of disease phase, and outcomes in 218 patients with myelofibrosis. *Cancer*. 2019.
14. Celgene. JAKARTA: A phase 3, multicenter, randomized, double-blind, placebo-controlled, 3-arm study of SAR302503 in patients with intermediate-2 or high-risk primary myelofibrosis, post-polycythemia vera myelofibrosis, or post-essential thrombocythemia myelofibrosis with splenomegaly (clinical study report). 13 Dec 2018 2018.
15. Mukuria C, Rowen D, Brazier JE, et al. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015; 18(6):846-55.
16. Rafia R and Mukuria CW. Exploratory Psychometric Analysis of The Eq-5d In A Myelofibrosis Population. *Value in Health*. 2015; 18(3):A211.

17. Alava MH, Wailoo AJ and Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value in Health*. 2012; 15(3):550-61.
18. Mesa RA, Kiladjian J-J, Verstovsek S, et al. Comparison of placebo and best available therapy for the treatment of myelofibrosis in the phase 3 COMFORT studies. *Haematologica*. 2014; 99(2):292-8.
19. Harrison CN, Mesa RA, Jamieson C, et al. Case series of potential Wernicke's encephalopathy in patients treated with Fedratinib. *Blood*. 2017; 130(Supplement 1):4197-.
20. Haematological Malignancy Research Network (HMRN). Haematological Malignancy Research Network: Clinical Management, Resource Utilisation and Outcome in Primary and Secondary Myelofibrosis. 27 May 2020 2020.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis ID1501

Follow up Clarification questions

August 2020

A1. Please provide in Excel format the following Kaplan-Meier (KM) survival functions and parameter estimates and variance-covariance matrices for each parametric distribution (using the format below). Please also provide for each of these outcomes (KM) how many patients had an event and how many were censored (using the table format below):

1. Time to treatment discontinuation [TTD] (including death as an event) for the overall JAKARTA-2 population with Int2/high risk (n=[redacted]) – from start of randomisation (including everyone – e.g patients affected [censored] or not by clinical hold),
2. TTD (including death as an event) for the overall JAKARTA-2 population with Int2/high risk (n=[redacted]) – from start of randomisation (removing patients affected by the clinical hold),
3. TTD (including death as an event) for responders (defined as spleen or symptoms) in JAKARTA-2 with Int2/high risk (n=[redacted]?) – from 24 weeks onward (similar to Figure 24 in CS, but death is not censored),
4. TTD (including death as an event) for non-responders (defined as spleen or symptoms) in JAKARTA-2 with Int2/high risk pop (n=[redacted]?) – from 24 weeks onward,
5. Time to death from any cause (OS) for responders (defined as spleen or symptoms) in JAKARTA-2 with Int2/high risk pop (n=[redacted]?) – from 24 weeks onward,
6. OS for non-responders (defined as spleen or symptoms) in JAKARTA-2 with Int2/high risk pop (n=[redacted]?) – from 24 weeks onward.

The requested outputs have been provided in an Excel document.

Sample Table for KM:

Product-Limit Survival Estimates					
Weeks	**Survival**	**Failure**	**Survival Standard Error**	**Number Event**	**Number Censored**
0	1	0	.	.	.
2	0.98	0.02	.	.	.
3	0.97	0.03	.	.	.
4	0.95	0.05	.	.	.
……	…..	…..	…..	…..	…..
1000	0.05	0.95	.	.	.
1001	0.02	0.98	.	.	.

Sample Table for number of events vs. censored:

	Total number of patient	Total number of events	Total number of censored events (death and any reason)	Number of patient censored due to death
TTD Int2/high (from randomisation)	.	.	.	0
TTD Int2/high Responders (from 24 weeks onward)	.	.	.	0
TTD Int2/high Non-Responders (from 24 weeks onward)	.	.	.	0
OS – Responders (from 24 weeks onward)	.	.	.	0
OS – Non-Responders (from 24 weeks onward)	.	.	.	0

Sample Table for parameters for parametric distribution (as set out in the economic model):

Exponential	Mean		Rate
Rate	.	Rate	.

Generalised gamma	Mean		Mu	Sigma	Q
Mu	.	Mu	.	.	.
Sigma	.	Sigma	.	.	.
Q	.	Q	.	.	.

Gompertz	Mean		Shape	Rate
Shape	.	Shape	.	.
Rate	.	Rate	.	.

Log-logistic	Mean		Shape	Scale
Shape	.	Shape	.	.
Scale	.	Scale	.	.

Log-normal	Mean		Meanlog	Sdlog
Meanlog	.	Meanlog	.	.
Sdlog	.	Sdlog	.	.

Weibull	Mean		Shape	Scale
Shape	.	Shape	.	.
Scale	.	Scale	.	.

Distribution	**AIC**	**BIC**
Exponential	.	.
Generalised gamma	.	.
Gompertz	.	.
Log-logistic	.	.
Log-normal	.	.
Weibull	.	.

A2. From the response to clarification question B4, it appears that death was censored when estimating TTD. Please provide the number of patients who were considered to have an event, censored due to death, and censored due to other reasons for the KM used for TTD (spleen and/or symptom) in the model for responders (Figure 24 in CS) and non-responders (Figure 23 in CS).

While the code for TTD was set up to censor for death after 24 weeks, upon investigation, it was found that no deaths were recorded which led to censoring for TTD after 24 weeks.

Therefore, the original concern of the ERG holds true that the TTD in the model predictions will be lower than the TTD parametric function.

The requested outputs have been provided in an Excel document.

Sample Table for events vs. censored:

	Total number of patient	Total number of events	Total number of censored events (death and any reason)	Number of patient censored due to death
TTD Responders	.	.	.	.
TTD Non-responder	.	.	.	.

A3. Thank you for your responses to clarification questions B10 for the ITT population (including Int1). Could you please provide (a) the KM for OS for the int2/high subgroup (n=[redacted]) in Excel format (using the format below) and (b) also provide the p-values:

- Other MF vs. PMF
- Intolerant vs. Resistant
- Relapsed vs. Refractory

The requested outputs have been provided in an Excel document.

Sample Table:

Product-Limit Survival Estimates					
Weeks	**Survival**	**Failure**	**Survival Standard Error**	**Number Event**	**Number Censored**
0	1	0	.	.	.
2	0.98	0.02	.	.	.
3	0.97	0.03	.	.	.
4	0.95	0.05	.	.	.
……	…..	…..	…..	…..	…..
1000	0.05	0.95	.	.	.
1001	0.02	0.98	.	.	.

A4. In response to clarification question A8, the company provided results for MAIC in Table 7. Can the company confirm that the point estimates and 95% CI risk difference is correct. In particular, confirm whether the lower CI should be negative i.e. should it be [redacted]

For this new analysis, there was a mistake in the bootstrapping code for the confidence interval of the risk difference for SVR. The results for SVR have now been corrected and updated in Table 1. A histogram of the bootstrap risk differences (as proportions) are included in Figure 1. This error was not made in the original analyses.

Table 1: Additional MAIC analyses for SVR to include all possible IPSS/DIPSS-PLUS predictors in matching (correction made)

Endpoint	Method	Variables included in adjustment	JAKARTA-2 (400 mg FEDR)	SIMPLIFY-2 (BAT)
SVR	MAIC	ECOG PS DIPSS Transfusion dependence Age	[redacted]% (CI:[redacted])	5.8% (n=3; N=52)
			Risk difference (FEDR versus BAT) [95% CI]: [redacted]	

Figure 1: Histogram of the bootstrap RDs for SVR

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis ID1501

Follow up Clarification questions

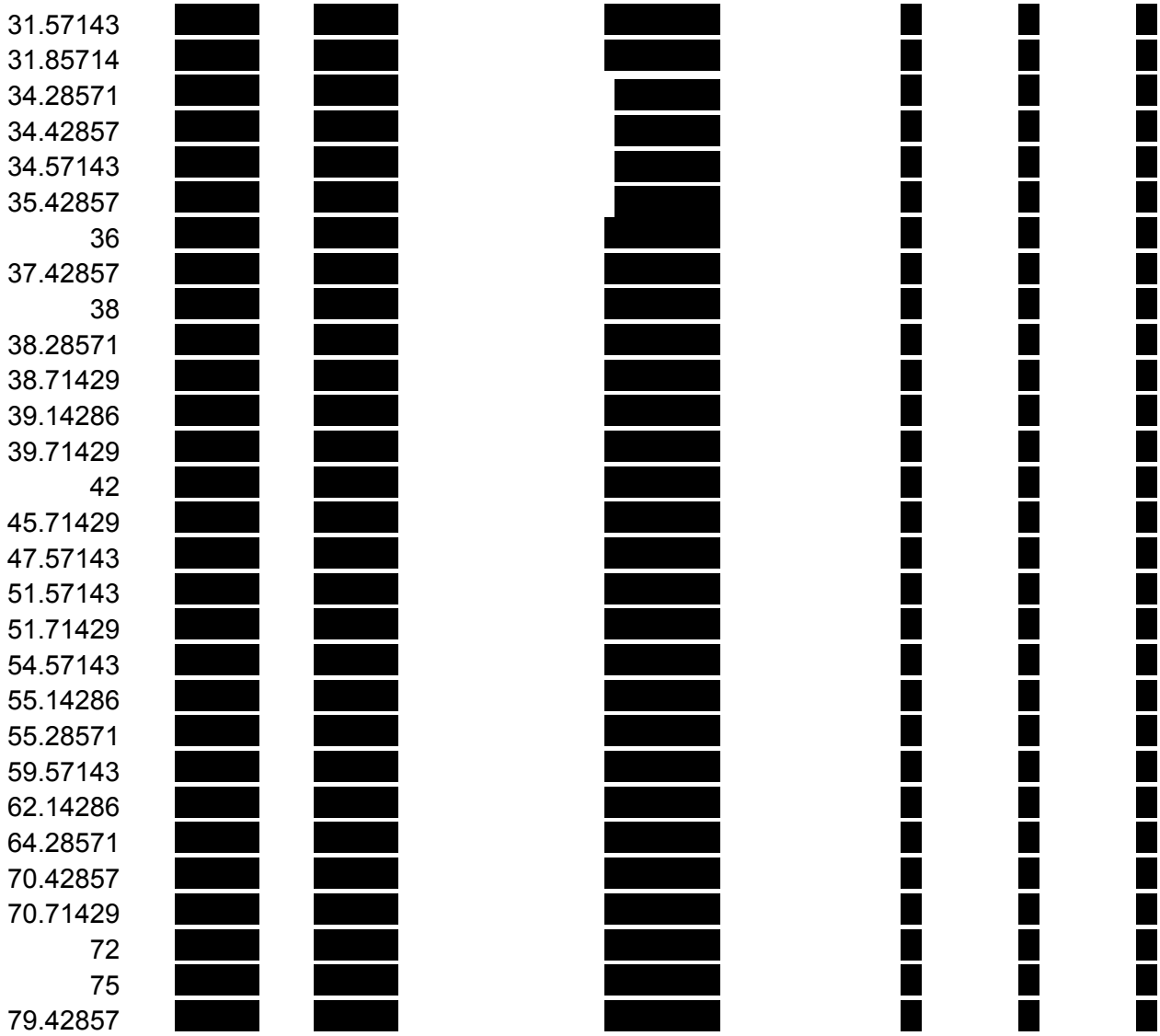
August 2020

A1 Summary Table

	Total.number. of.patients	Total.number. of.events	Total.number. of.censors	Number of patient censored due to death	Censor.criteria
A1.1	[redacted]	[redacted]	[redacted]	[redacted]	dcsreas == 'STUDY TERMINATED BY SPONSOR'
A1.2	[redacted]	[redacted]	[redacted]	[redacted]	dcsreas == 'STUDY TERMINATED BY SPONSOR' (Those affected by clinical hold' was interpreted as being terminated before 24 weeks, those terminated after 24 weeks were still included)
A1.3	[redacted]	[redacted]	[redacted]	[redacted]	dcsreas == 'STUDY TERMINATED BY SPONSOR'
A1.4	[redacted]	[redacted]	[redacted]	[redacted]	dcsreas == 'STUDY TERMINATED BY SPONSOR'
A1.5	[redacted]	[redacted]	[redacted]	[redacted]	CNSR = 0 (OS censor taken from ados)
A1.6	[redacted]	[redacted]	[redacted]	[redacted]	CNSR = 0 (OS censor taken from ados)

A1_1

Weeks	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
0.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2.857143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.142857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.285714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
9.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
11.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
12.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
13.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
16.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
16.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
16.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
18.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
19	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
19.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
19.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
20.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
21.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
21.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
22	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
22.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
26.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
26.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
27.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]


31.57143
31.85714
34.28571
34.42857
34.57143
35.42857
36
37.42857
38
38.28571
38.71429
39.14286
39.71429
42
45.71429
47.57143
51.57143
51.71429
54.57143
55.14286
55.28571
59.57143
62.14286
64.28571
70.42857
70.71429
72
75
79.42857

A1_1

Exponential	Mean	Rate
Rate	[redacted]	[redacted]

Generalised gamma	Mean	Mu	Sigma	Q
Mu	[redacted]	[redacted]	[redacted]	[redacted]
Sigma	[redacted]	[redacted]	[redacted]	[redacted]
Q	[redacted]	[redacted]	[redacted]	[redacted]

Gompertz	Mean	Shape	Rate
Shape	[redacted]	[redacted]	[redacted]
Rate	[redacted]	[redacted]	[redacted]

Log-logistic	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Log-normal	Mean	Meanlog	Sdlog
Meanlog	[redacted]	[redacted]	[redacted]
Sdlog	[redacted]	[redacted]	[redacted]

Weibull	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Distribution	**AIC**	**BIC**
Exponential	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]
Weibull	[redacted]	[redacted]

A1_2

Weeks	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
0.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2.857143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.142857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.285714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
9.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
11.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
16.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
18.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
19	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
20.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
21.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
21.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
22	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
26.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
26.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
27.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
34.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
34.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
34.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
35.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
36	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
37.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

38	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
38.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
38.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
42	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
45.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
47.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
51.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
51.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
54.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
55.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
55.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
59.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
62.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
64.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
70.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
70.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
72	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
75	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
79.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

A1_2

Exponential

	Mean	Rate
Rate	[redacted]	[redacted]

Generalised gamma

	Mean	Mu	Sigma	Q
Mu	[redacted]	[redacted]	[redacted]	[redacted]
Sigma	[redacted]	[redacted]	[redacted]	[redacted]
Q	[redacted]	[redacted]	[redacted]	[redacted]

Gompertz

	Mean	Shape	Rate
Shape	[redacted]	[redacted]	[redacted]
Rate	[redacted]	[redacted]	[redacted]

Log-logistic

	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Log-normal

	Mean	Meanlog	Sdlog
Meanlog	[redacted]	[redacted]	[redacted]
Sdlog	[redacted]	[redacted]	[redacted]

Weibull

	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Distribution	**AIC**	**BIC**
Exponential	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]

Log-logistic	[redacted]	[redacted]	
Log-normal	[redacted]	[redacted]	
Weibull	[redacted]	[redacted]	

A1_3

Weeks	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
0.285714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1.142857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
7.428571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
7.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
10.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
10.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
10.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
11.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
13.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
14	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
14.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
18	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
21.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
23.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
27.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
27.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
35.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
38.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
46.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
46.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
48	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
51	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
55.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

A1_3

Exponential	Mean	Rate
Rate	[redacted]	[redacted]

Generalised gamma	Mean	Mu	Sigma	Q
Mu	[redacted]	[redacted]	[redacted]	[redacted]
Sigma	[redacted]	[redacted]	[redacted]	[redacted]
Q	[redacted]	[redacted]	[redacted]	[redacted]

Gompertz	Mean	Shape	Rate
Shape	[redacted]	[redacted]	[redacted]
Rate	[redacted]	[redacted]	[redacted]

Log-logistic	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Log-normal	Mean	Meanlog	Sdlog
Meanlog	[redacted]	[redacted]	[redacted]
Sdlog	[redacted]	[redacted]	[redacted]

Weibull	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Distribution	**AIC**	**BIC**
Exponential	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]
Weibull	[redacted]	[redacted]

A1_4

Weeks	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
3.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
7.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
7.857143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
14.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
40.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

A1_4

Exponential	Mean	Rate
Rate	[redacted]	[redacted]

Generalised gamma	Mean	Mu	Sigma	Q
Mu	[redacted]	[redacted]	[redacted]	[redacted]
Sigma	[redacted]	[redacted]	[redacted]	[redacted]
Q	[redacted]	[redacted]	[redacted]	[redacted]

Gompertz	Mean	Shape	Rate
Shape	[redacted]	[redacted]	[redacted]
Rate	[redacted]	[redacted]	[redacted]

Log-logistic	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Log-normal	Mean	Meanlog	Sdlog
Meanlog	[redacted]	[redacted]	[redacted]
Sdlog	[redacted]	[redacted]	[redacted]

Weibull	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Distribution	**AIC**	**BIC**
Exponential	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]
Weibull	[redacted]	[redacted]

A1_5

Weeks	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
9.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
14.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
16	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
17.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
19.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
21.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
23.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
26.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
27.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
28.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
28.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
31.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
34.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
37.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
38.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
40.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
42.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
44.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
44.57143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
46.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
52.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
56.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
61.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
63.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
64.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
67	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
68	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

A1_5

Exponential	Mean	Rate
Rate	[redacted]	[redacted]

Generalised gamma	Mean	Mu	Sigma	Q
Mu	[redacted]	[redacted]	[redacted]	[redacted]
Sigma	[redacted]	[redacted]	[redacted]	[redacted]
Q	[redacted]	[redacted]	[redacted]	[redacted]

Gompertz	Mean	Shape	Rate
Shape	[redacted]	[redacted]	[redacted]
Rate	[redacted]	[redacted]	[redacted]

Log-logistic	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Log-normal	Mean	Meanlog	Sdlog
Meanlog	[redacted]	[redacted]	[redacted]
Sdlog	[redacted]	[redacted]	[redacted]

Weibull	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Distribution	**AIC**	**BIC**
Exponential	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]
Weibull	[redacted]	[redacted]

A1_6

Weeks	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
0.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2.857143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3.857143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5.285714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6.714286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
8.571429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
9.142857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
9.857143	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
11.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
12.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
12.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
16.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
22.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
23.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
24.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
26.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
29.71429	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
33.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
41.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
45.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
46.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
48	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
49.85714	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
50.42857	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
51.28571	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
61.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
65.14286	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

A1_6

Exponential	Mean	Rate
Rate	[redacted]	[redacted]

Generalised gamma	Mean	Mu	Sigma	Q
Mu	[redacted]	[redacted]	[redacted]	[redacted]
Sigma	[redacted]	[redacted]	[redacted]	[redacted]
Q	[redacted]	[redacted]	[redacted]	[redacted]

Gompertz	Mean	Shape	Rate
Shape	[redacted]	[redacted]	[redacted]
Rate	[redacted]	[redacted]	[redacted]

Log-logistic	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Log-normal	Mean	Meanlog	Sdlog
Meanlog	[redacted]	[redacted]	[redacted]
Sdlog	[redacted]	[redacted]	[redacted]

Weibull	Mean	Shape	Scale
Shape	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]

Distribution	**AIC**	**BIC**
Exponential	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]
Weibull	[redacted]	[redacted]

A2 Summary Table

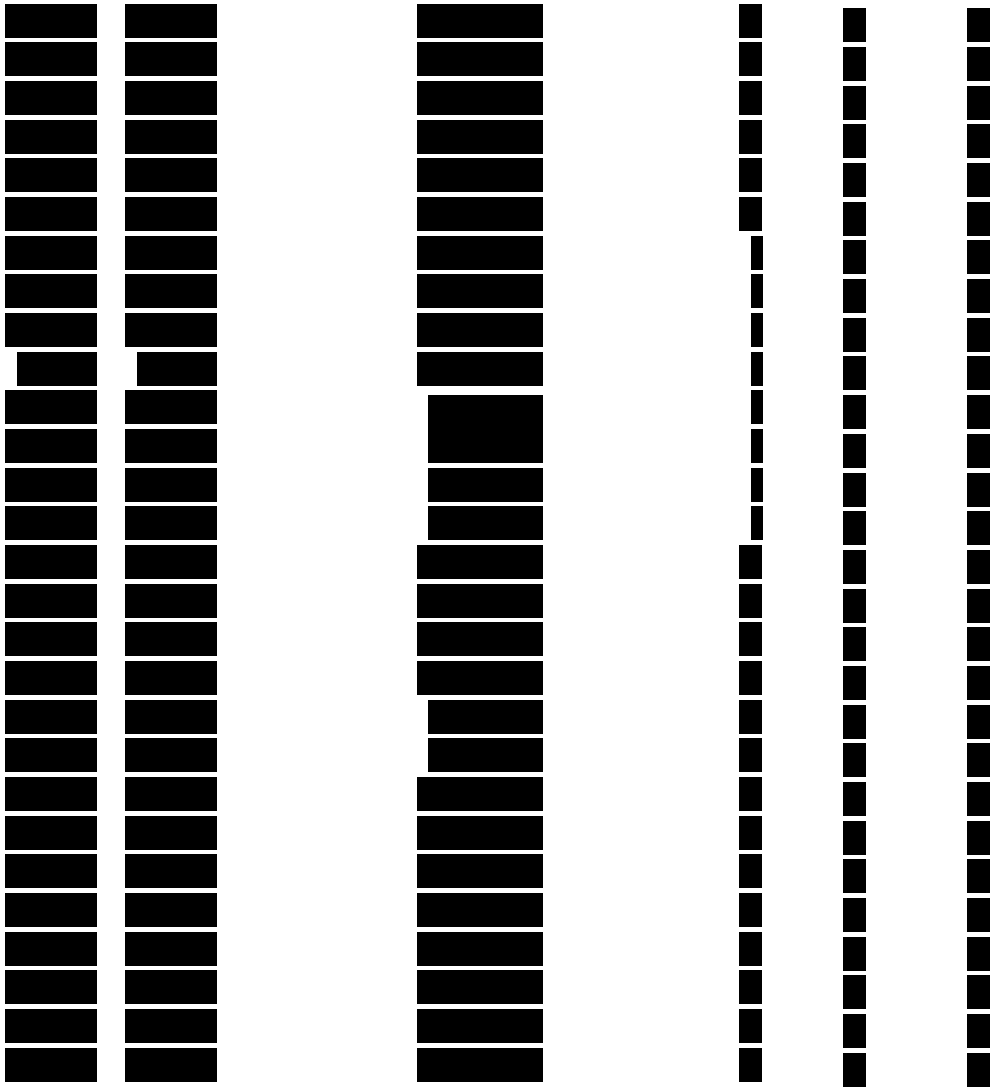
	Total number of patients	Total number of events	Total number of censored events	Total number of patients censored for death
TTD responders (any response)	[redacted]	[redacted]	[redacted]	[redacted]
TTD non-responders (any response)	[redacted]	[redacted]	[redacted]	[redacted]
TTD non-responders (SVR response)	[redacted]	[redacted]	[redacted]	[redacted]

A3_MF

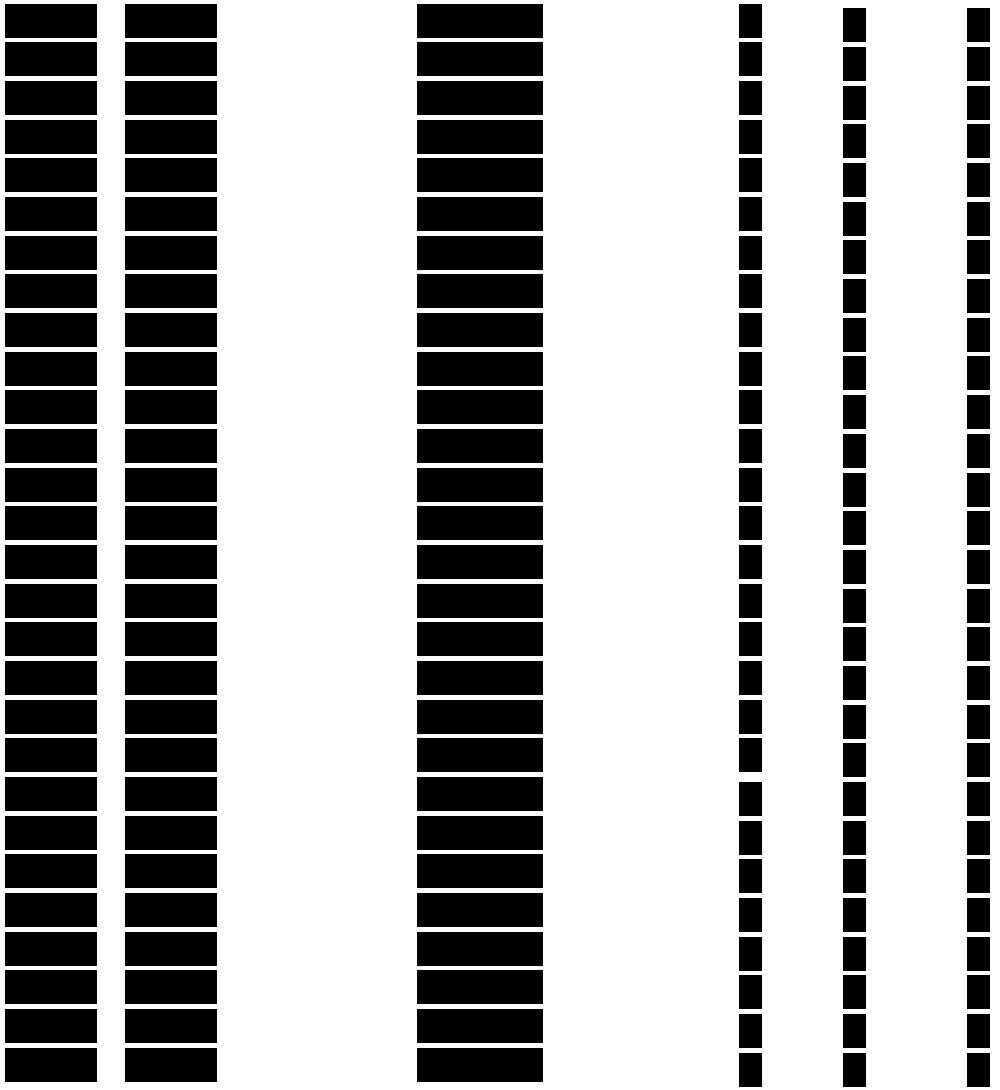
	x
1	p-value: 0.498911801847692

	Weeks	Var	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
1	0.714286	Var=Other myelofibrosis type						
2	11.85714	Var=Other myelofibrosis type						
3	23.71429	Var=Other myelofibrosis type						
4	24.71429	Var=Other myelofibrosis type						
5	26.85714	Var=Other myelofibrosis type						
6	29.28571	Var=Other myelofibrosis type						
7	33.14286	Var=Other myelofibrosis type						
8	33.71429	Var=Other myelofibrosis type						
9	35.14286	Var=Other myelofibrosis type						
10	36.42857	Var=Other myelofibrosis type						
11	39.14286	Var=Other myelofibrosis type						
12	41.14286	Var=Other myelofibrosis type						
13	45.71429	Var=Other myelofibrosis type						
14	47.28571	Var=Other myelofibrosis type						
15	48.14286	Var=Other myelofibrosis type						
16	48.42857	Var=Other myelofibrosis type						
17	48.57143	Var=Other myelofibrosis type						
18	50.14286	Var=Other myelofibrosis type						
19	51.14286	Var=Other myelofibrosis type						
20	52.14286	Var=Other myelofibrosis type						
21	52.71429	Var=Other myelofibrosis type						

22	53.14286	Var=Other myelofibrosis type
23	55.28571	Var=Other myelofibrosis type
24	57.14286	Var=Other myelofibrosis type
25	58.14286	Var=Other myelofibrosis type
26	61.42857	Var=Other myelofibrosis type
27	64.42857	Var=Other myelofibrosis type
28	65.14286	Var=Other myelofibrosis type
29	68.14286	Var=Other myelofibrosis type
30	70.85714	Var=Other myelofibrosis type
31	73.85714	Var=Other myelofibrosis type
32	74.42857	Var=Other myelofibrosis type
33	76.14286	Var=Other myelofibrosis type
34	80.85714	Var=Other myelofibrosis type
35	87.14286	Var=Other myelofibrosis type
36	3.285714	Var=Primary myelofibrosis
37	6.142857	Var=Primary myelofibrosis
38	12	Var=Primary myelofibrosis
39	12.42857	Var=Primary myelofibrosis
40	19.42857	Var=Primary myelofibrosis
41	20.14286	Var=Primary myelofibrosis
42	22.28571	Var=Primary myelofibrosis
43	27.71429	Var=Primary myelofibrosis
44	27.85714	Var=Primary myelofibrosis
45	30	Var=Primary myelofibrosis
46	30.71429	Var=Primary myelofibrosis
47	32.57143	Var=Primary myelofibrosis
48	33.85714	Var=Primary myelofibrosis
49	36.14286	Var=Primary myelofibrosis



50	38.28571	Var=Primary myelofibrosis
51	40	Var=Primary myelofibrosis
52	40.71429	Var=Primary myelofibrosis
53	41.14286	Var=Primary myelofibrosis
54	43.57143	Var=Primary myelofibrosis
55	44	Var=Primary myelofibrosis
56	46.85714	Var=Primary myelofibrosis
57	47.14286	Var=Primary myelofibrosis
58	48.71429	Var=Primary myelofibrosis
59	49.71429	Var=Primary myelofibrosis
60	49.85714	Var=Primary myelofibrosis
61	50.57143	Var=Primary myelofibrosis
62	53.14286	Var=Primary myelofibrosis
63	53.71429	Var=Primary myelofibrosis
64	62.71429	Var=Primary myelofibrosis
65	63.57143	Var=Primary myelofibrosis
66	64.42857	Var=Primary myelofibrosis
67	66.57143	Var=Primary myelofibrosis
68	68.57143	Var=Primary myelofibrosis
69	69.28571	Var=Primary myelofibrosis
70	70.71429	Var=Primary myelofibrosis
71	72	Var=Primary myelofibrosis
72	75.28571	Var=Primary myelofibrosis
73	85.14286	Var=Primary myelofibrosis
74	85.42857	Var=Primary myelofibrosis
75	88.14286	Var=Primary myelofibrosis
76	89.14286	Var=Primary myelofibrosis
77	91	Var=Primary myelofibrosis



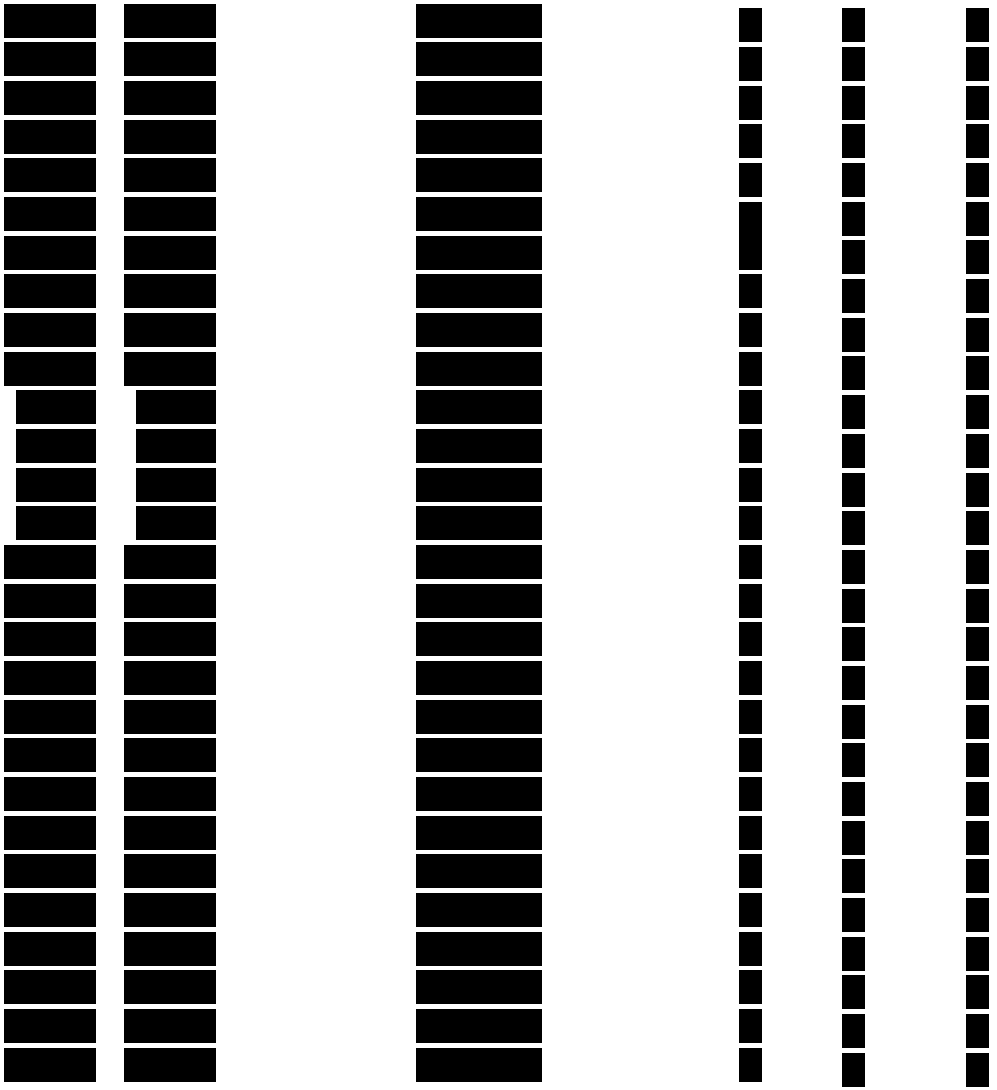
78	92	Var=Primary myelofibrosis	[redacted]	[redacted]		[redacted]	[redacted]	[redacted]

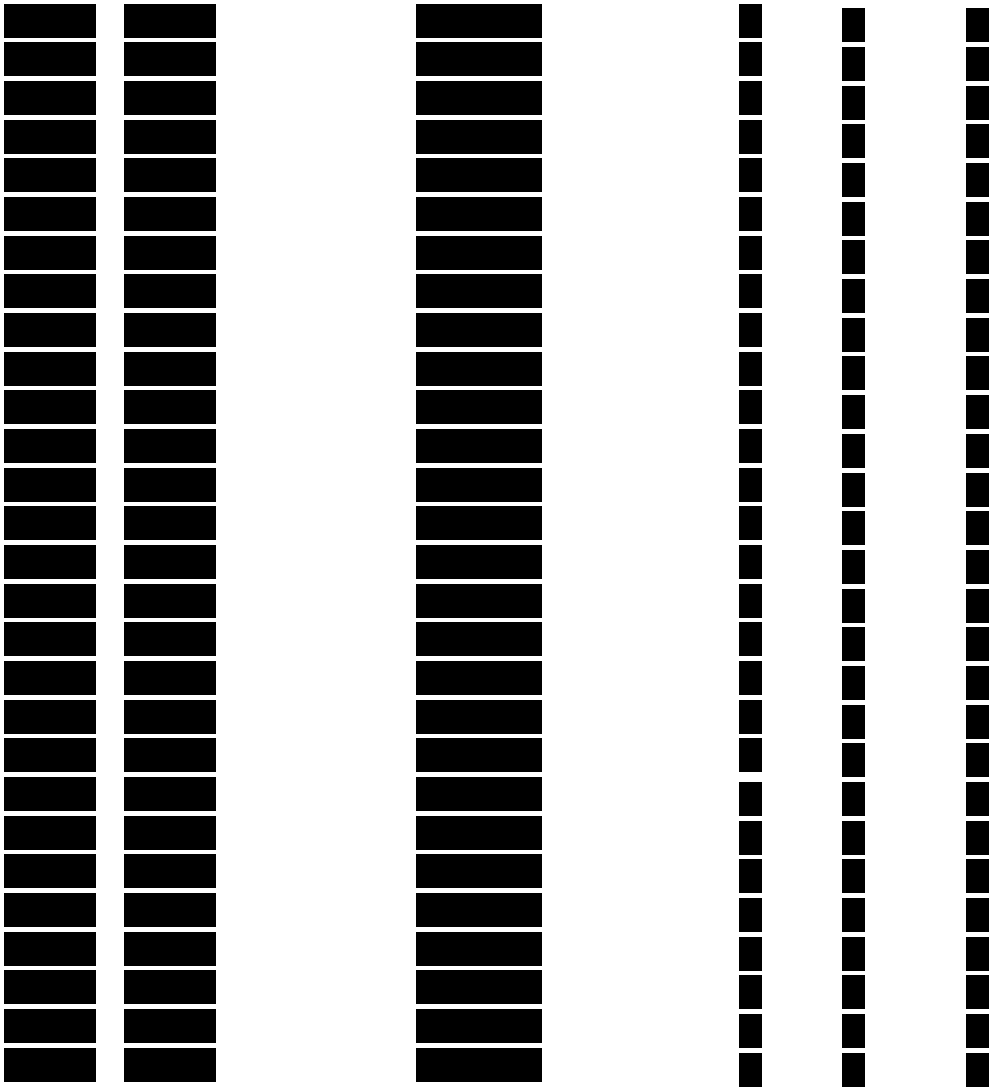
A3_Res

	x
1	p-value: 0.230626240184595

	Weeks	Var	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
1	0.714286	Var=Intolerant						
2	3.285714	Var=Intolerant						
3	11.85714	Var=Intolerant						
4	12	Var=Intolerant						
5	19.42857	Var=Intolerant						
6	23.71429	Var=Intolerant						
7	29.28571	Var=Intolerant						
8	33.71429	Var=Intolerant						
9	36.14286	Var=Intolerant						
10	36.42857	Var=Intolerant						
11	38.28571	Var=Intolerant						
12	39.14286	Var=Intolerant						
13	40	Var=Intolerant						
14	41.14286	Var=Intolerant						
15	44	Var=Intolerant						
16	47.14286	Var=Intolerant						
17	51.14286	Var=Intolerant						
18	57.14286	Var=Intolerant						
19	61.42857	Var=Intolerant						
20	64.42857	Var=Intolerant						
21	68.57143	Var=Intolerant						

22	70.85714	Var=Intolerant
23	73.85714	Var=Intolerant
24	74.42857	Var=Intolerant
25	80.85714	Var=Intolerant
26	85.42857	Var=Intolerant
27	91	Var=Intolerant
28	6.142857	Var=Resistant
29	12.42857	Var=Resistant
30	20.14286	Var=Resistant
31	22.28571	Var=Resistant
32	24.71429	Var=Resistant
33	26.85714	Var=Resistant
34	27.71429	Var=Resistant
35	27.85714	Var=Resistant
36	30	Var=Resistant
37	30.71429	Var=Resistant
38	32.57143	Var=Resistant
39	33.14286	Var=Resistant
40	33.85714	Var=Resistant
41	35.14286	Var=Resistant
42	40.71429	Var=Resistant
43	41.14286	Var=Resistant
44	43.57143	Var=Resistant
45	45.71429	Var=Resistant
46	46.85714	Var=Resistant
47	47.14286	Var=Resistant
48	47.28571	Var=Resistant
49	48.14286	Var=Resistant



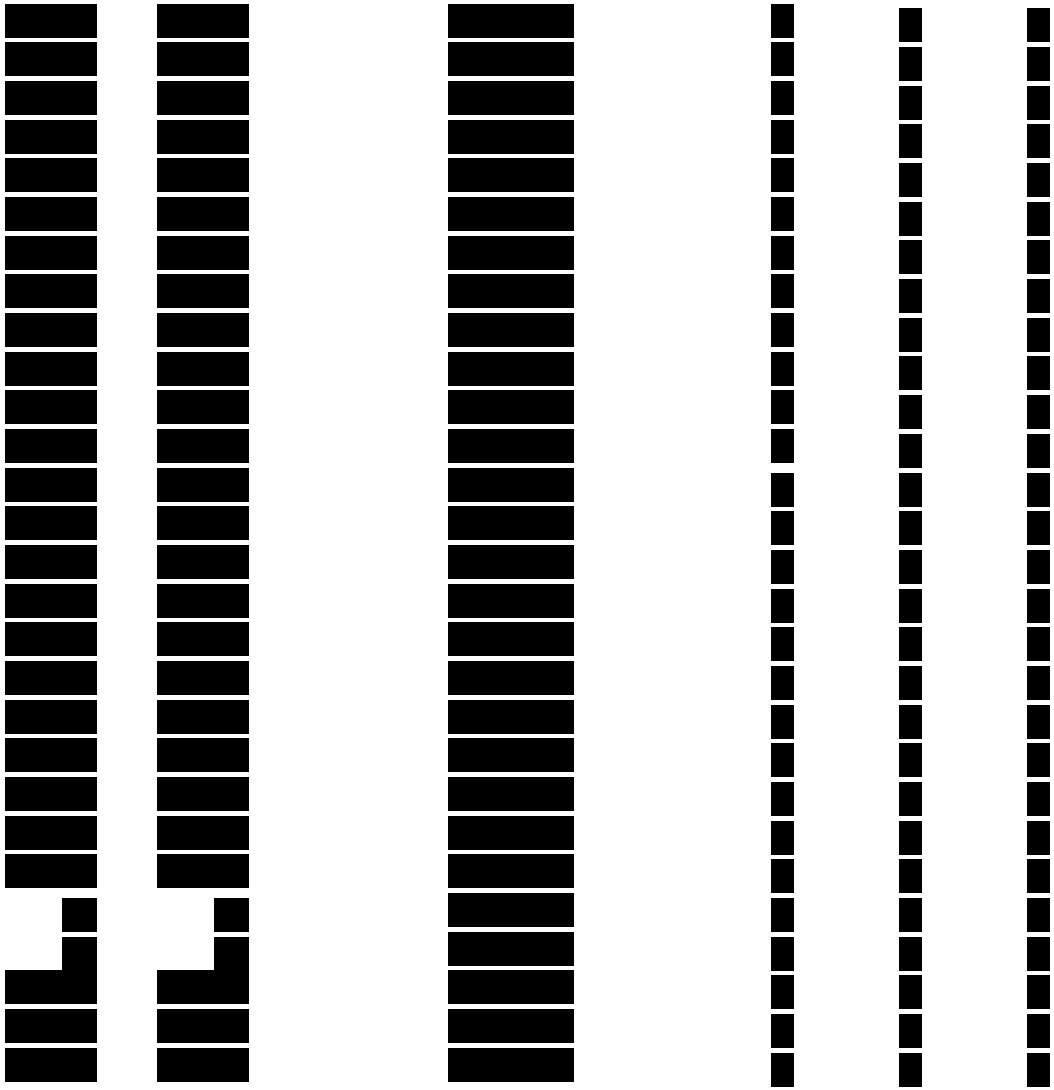
50	48.42857	Var=Resistant	
51	48.57143	Var=Resistant	
52	48.71429	Var=Resistant	
53	49.71429	Var=Resistant	
54	49.85714	Var=Resistant	
55	50.14286	Var=Resistant	
56	50.57143	Var=Resistant	
57	52.14286	Var=Resistant	
58	52.71429	Var=Resistant	
59	53.14286	Var=Resistant	
60	53.71429	Var=Resistant	
61	55.28571	Var=Resistant	
62	58.14286	Var=Resistant	
63	62.71429	Var=Resistant	
64	64.42857	Var=Resistant	
65	65.14286	Var=Resistant	
66	66.57143	Var=Resistant	
67	68.14286	Var=Resistant	
68	69.28571	Var=Resistant	
69	70.71429	Var=Resistant	
70	72	Var=Resistant	
71	75.28571	Var=Resistant	
72	76.14286	Var=Resistant	
73	85.14286	Var=Resistant	
74	87.14286	Var=Resistant	
75	88.14286	Var=Resistant	
76	89.14286	Var=Resistant	
77	92	Var=Resistant	

A3_RR

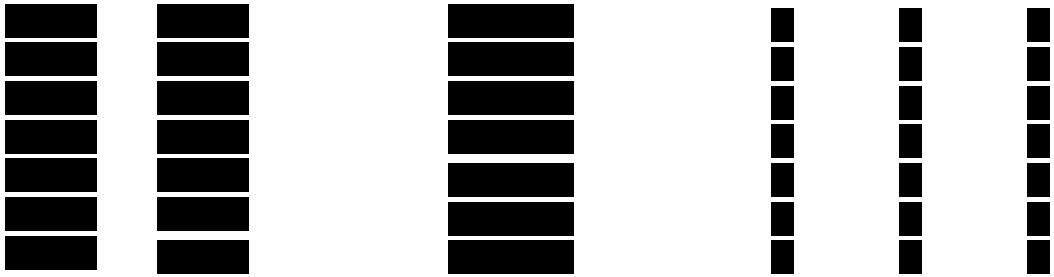
	x
	p-value:
1	0.166667612397634

	Weeks	Var	Survival	Failure	Survival.standard.error	Number.at.risk	Event	Censor
1	12	Var=Refractory						
2	12.42857	Var=Refractory						
3	20.14286	Var=Refractory						
4	23.71429	Var=Refractory						
5	27.71429	Var=Refractory						
6	27.85714	Var=Refractory						
7	30	Var=Refractory						
8	32.57143	Var=Refractory						
9	33.71429	Var=Refractory						
10	33.85714	Var=Refractory						
11	35.14286	Var=Refractory						
12	36.42857	Var=Refractory						
13	39.14286	Var=Refractory						
14	40	Var=Refractory						
15	40.71429	Var=Refractory						
16	41.14286	Var=Refractory						
17	45.71429	Var=Refractory						
18	47.14286	Var=Refractory						
19	48.14286	Var=Refractory						
20	49.71429	Var=Refractory						

21	49.85714	Var=Refractory
22	52.14286	Var=Refractory
23	52.71429	Var=Refractory
24	53.14286	Var=Refractory
25	53.71429	Var=Refractory
26	55.28571	Var=Refractory
27	58.14286	Var=Refractory
28	62.71429	Var=Refractory
29	63.57143	Var=Refractory
30	64.42857	Var=Refractory
31	66.57143	Var=Refractory
32	69.28571	Var=Refractory
33	70.71429	Var=Refractory
34	72	Var=Refractory
35	73.85714	Var=Refractory
36	75.28571	Var=Refractory
37	80.85714	Var=Refractory
38	87.14286	Var=Refractory
39	88.14286	Var=Refractory
40	89.14286	Var=Refractory
41	91	Var=Refractory
42	11.85714	Var=Relapsed
43	19.42857	Var=Relapsed
44	22.28571	Var=Relapsed
45	30.71429	Var=Relapsed
46	43.57143	Var=Relapsed
47	44	Var=Relapsed
48	46.85714	Var=Relapsed



49	50.14286	Var=Relapsed
50	50.57143	Var=Relapsed
51	53.14286	Var=Relapsed
52	61.42857	Var=Relapsed
53	68.57143	Var=Relapsed
54	70.85714	Var=Relapsed
55	92	Var=Relapsed



NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis ID1501

Addendum clarification questions

March 2021

File name	Version	Contains confidential information	Date
ERG clarification post company addendum	**0.4**	**Yes**	**26/03/2021**

Notes for company

Highlighting in the template

Square brackets and grey highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [grey highlighting] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

Section A: Clarification on effectiveness data

There are no effectiveness clarification questions on the addendum submission.

Section B: Clarification on cost-effectiveness data

Model features

B1. The ERG conducted a series of simple tests to check the model's implementation. Inconsistencies were identified raising concerns regarding the general implementation of the model. One of the tests carried out by the ERG was to check that the model produces the same QALYs for each arm under the same

assumptions. The following changes were made to remove any confounding impact associated with response assessment:

- Controls Sheet - Set Cell J24 (c_override) to “Equal OS and time on treatment”
- Controls Sheet - Set Cell J79 (c_manual_rux_2) to = c_manual_rux
- Controls Sheet – Set Cell J124 (c_utility_gender_text) to “No”
- Controls Sheet - Set Cell J128 (c_include_disutility_text) to “No”
- Replace in VB, “If Week_Looper <= 4 Then” TO “If Week_Looper <= 0 Then”

The following incremental QALYs are predicted using the same response rates as in the table below (before and after correcting for the inappropriate change made to utility value for BAT non-responders [assumption of no increment]).

	Test1 • FED=[redacted] • BAT=[redacted]	Test2 • FED=10% • BAT=10%	Test3 • FED=90% • BAT=90%
Prior to correcting for the inappropriate assumption of no increment in utility for non-responders for patients initiated on BAT			
Incremental QALYs	-0.010	0.035	-0.079
After correcting for the inappropriate assumption of no increment in utility for non-responders for patients initiated on BAT Utility sheet – Set Cell E30:F30 to Cell E28:F28			
Incremental QALYs	-0.020	-0.005	-0.043

In addition to the tests described above, the model generates inappropriate QALYs when setting utility increments to be the same irrespective of response status. The same changes as above were made to remove any confounding impact associated with response assessment. The following tests were then conducted:

	Test1 • Cell E27:F30 (all increments) = 0	Test2 • Cell E27:F30 (all increments) = +0.10
Incremental QALYs	0.0002	0.0002

Please check the model’s implementation in relation to these inconsistencies.

ERG tests (1)

To remove any confounding impact associated with response assessment, the ERG needed to make the following further model changes.

1. The ERG aimed to apply utility increments from week 0. Therefore, the following VB logic must be amended for these tests. From:

```
Baseline = Utility_Array(1, Sex)

Inc_Response = Utility_Array(2 - 2 * (State = State_BAT), Sex)
Inc_Non_Response = Utility_Array(3 - 2 * (State = State_BAT), Sex)

If Total_Weeks < 1 Then
    Func_Utility_Tx = Baseline * Func_Discount(Start, Start + Total_Weeks, -Discounted * D_Rate,
True)
Else
```

To:

```
Baseline = Utility_Array(1, Sex)

Inc_Response = Utility_Array(2 - 2 * (State = State_BAT), Sex)
Inc_Non_Response = Utility_Array(3 - 2 * (State = State_BAT), Sex)

If Response = 1 Then
    Utility_Tracker = Baseline + Inc_Response
Else
    Utility_Tracker = Baseline + Inc_Non_Response
End If

If Total_Weeks < 1 Then
    Func_Utility_Tx = Utility_Tracker * Func_Discount(Start, Start + Total_Weeks, D_Rate, True)
Else
```

Please note that *"-Discounted *"* should be removed as this was not updated during technical engagement. There is no impact on results to the decimal places shown in the table below.

2. A second change for these tests must be made to reflect that the company model does not allow patients who initiated on fedratinib to respond to BAT. Therefore, the following lines in VB should be removed for these tests:

```
'Assume no second response assessment
OUT_BAT_responder = 0
```

After making these changes, the following incremental QALYs are predicted:

	Test1 • FED=[redacted] • BAT=[redacted]	Test2 • FED=10% • BAT=10%	Test3 • FED=90% • BAT=90%
Prior to removing assumption of no increment in utility for non-responders for patients on BAT			
Incremental QALYs	0.030	0.045	0.006
After removing assumption of no increment in utility for non-responders for patients on BAT Utility sheet – Set Cell E30:F30 to Cell E28:F28			
Incremental QALYs	0.000	0.000	0.000

ERG tests (2)

Following the above amends, some minor differences (beyond 3 decimal places) remained in discounted QALYs. It was identified that all remaining differences in discounted QALYs were explained by the application of age-related utility adjustment, which did not account for changes in age discounted by the QALY discount rate.

To correct this, the following code can be added to declare the relevant variables:

```
Public OUT_LYs_DQ() As Double ' LYs (discounted by QALY discount rate)
Public OUT_LYs_DQ_Cum() As Double ' Cumulative LYs (discounted by QALY discount rate)
```

The following code can be added to ensure the relevant arrays are reset:

```
ReDim OUT_LYs_DQ(1 To NUM_Patients, 1 To 3)
ReDim OUT_LYs_DQ_Cum(1 To NUM_Patients, 1 To 3)
```

The following code can be added to track the relevant outcomes at the end of SUB_LYs():

```
OUT_LYs_DQ(Patient, State_JAK) = Func_Discount(0, OUT_LYs_Cum(Patient, State_JAK), DR_QALYs, False)
OUT_LYs_DQ(Patient, State_BAT) = Func_Discount(OUT_LYs_Cum(Patient, State_JAK),
OUT_LYs_Cum(Patient, State_BAT), DR_QALYs, False)
OUT_LYs_DQ(Patient, State_Care) = Func_Discount(OUT_LYs_Cum(Patient, State_BAT),
OUT_LYs_Cum(Patient, State_Care), DR_QALYs, False)

OUT_LYs_DQ_Cum(Patient, State_JAK) = OUT_LYs_DQ(Patient, State_JAK)
OUT_LYs_DQ_Cum(Patient, State_BAT) = OUT_LYs_DQ(Patient, State_JAK) + OUT_LYs_DQ(Patient,
State_BAT)
OUT_LYs_DQ_Cum(Patient, State_Care) = OUT_LYs_DQ(Patient, State_JAK) + OUT_LYs_DQ(Patient,
State_BAT) + OUT_LYs_DQ(Patient, State_Care)
```

The following code can be replaced. From:

```
If Include_Age_Utility Then
    OUT_QALYs_D(Patient, State_JAK) = OUT_QALYs_D(Patient, 1) * Func_Age_Utility_Adjustment(0,
OUT_LYs_Cum(Patient, 1), Age_Utility_Array, Sex_Male)
    OUT_QALYs_D(Patient, State_BAT) = OUT_QALYs_D(Patient, 2) *
Func_Age_Utility_Adjustment(OUT_LYs_Cum(Patient, 1), OUT_LYs_Cum(Patient, 2),
Age_Utility_Array, Sex_Male)
    OUT_QALYs_D(Patient, State_Care) = OUT_QALYs_D(Patient, 3) *
Func_Age_Utility_Adjustment(OUT_LYs_Cum(Patient, 2), OUT_LYs_Cum(Patient, 3),
Age_Utility_Array, Sex_Male)
End If
```

To:

```
If Include_Age_Utility Then
    OUT_QALYs_D(Patient, State_JAK) = OUT_QALYs_D(Patient, 1) * Func_Age_Utility_Adjustment(0,
OUT_LYs_DQ_Cum(Patient, 1), Age_Utility_Array, Sex_Male)
    OUT_QALYs_D(Patient, State_BAT) = OUT_QALYs_D(Patient, 2) *
Func_Age_Utility_Adjustment(OUT_LYs_DQ_Cum(Patient, 1), OUT_LYs_DQ_Cum(Patient, 2),
Age_Utility_Array, Sex_Male)
    OUT_QALYs_D(Patient, State_Care) = OUT_QALYs_D(Patient, 3) *
Func_Age_Utility_Adjustment(OUT_LYs_DQ_Cum(Patient, 2), OUT_LYs_DQ_Cum(Patient, 3),
Age_Utility_Array, Sex_Male)
End If
```

The ERG's second set of testing (which involved setting utility increments to be the same irrespective of response status) with the above amends produced the following results:

	Test1 • Cell E27:F30 (all increments) = 0	Test2 • Cell E27:F30 (all increments) = +0.10
Prior to amending age utility adjustment for discounted QALYs		
Incremental QALYs	0.0002	0.0001
After amending age utility adjustment for discounted QALYs		
Incremental QALYs	0.0000 (to all decimal places)	0.0000 (to all decimal places)

Overall, the impact of such corrections on the company base case ICER are very small:

	ICER	**Change from company base case**
Company base case	£24,784.02	-
Correction to remove "-Discounted *" code	£24,784.20	+ £0.18
Correction to age utility adjustment	£24,735,94	- £48.08

The ERG describes 'correcting for the inappropriate assumption of no increment in utility for non-responders for patients initiated on BAT'. The company maintain that this assumption is appropriate and further clarification is provided in the company response to B2. We also would like to highlight that patients are not being initiated on BAT, they are continuing on BAT.

Utilities

B2. Please clarify why the increment in utility values for non-responders is decreased [redacted] compared with the original response definition [redacted], when responders (who got up-titrated) are included in the non-responder group. Please discuss any implications for the BAT arm (when utility increments are used for non-responders).

For the addendum, utility analyses were updated with the new definition of response (at end of cycle 6) but were also further updated to include all available post-baseline utility values (rather than only those at end of cycle 3 and end of cycle 6). Therefore the updated analyses are not directly comparable to the original analyses.

The change was considered necessary because the new response definition data were only available at EOC6, and so the analysis was split by response at that time point. It was also felt that this analysis would better utilize all of the available data and align closer with the modelling (response at the end of cycle 3 was not of interest). The results are presented in the table below

Responders	Up-titrated	Not up-titrated
Number of patients	[redacted]	[redacted]
Number of observations	[redacted]	[redacted]
Mean MF-8D utility	[redacted]	[redacted]
Standard deviation MF-8D utility	[redacted]	[redacted]
Median MF-8D utility	[redacted]	[redacted]
Minimum MF-8D utility	[redacted]	[redacted]
Maximum MF-8D utility	[redacted]	[redacted]

The table indicates that average utility was relatively consistent across the two groups (responders up-titrated and not up-titrated), therefore it is likely that the change in results was due to the update in the methods rather than the original expectation that responders who were up-titrated had worse utility values than those who were not up-titrated.

When utility increments are used for non-responders in the model (as a scenario), the MF-8D utility data from the fedratinib arm of JAKARTA-2 is used for both arms. Yet, there are difference in total symptom score (TSS) changes observed between the non-responders in JAKARTA-2 (on fedratinib) and SIMPLIFY-2 (on BAT) as seen in Figure 1 and Figure 2.

Figure 1: SIMPLIFY-2 TSS results (Harrison et al. 2018)[1]

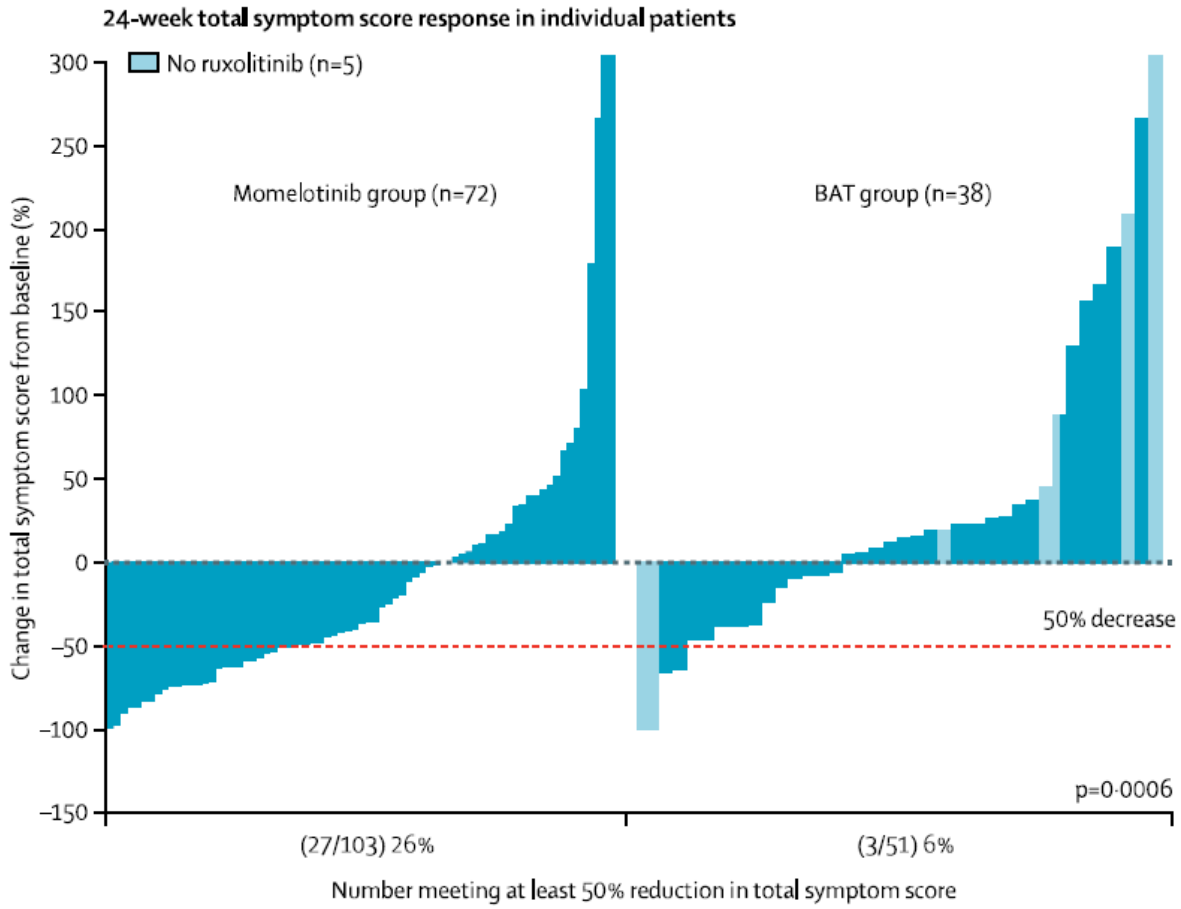
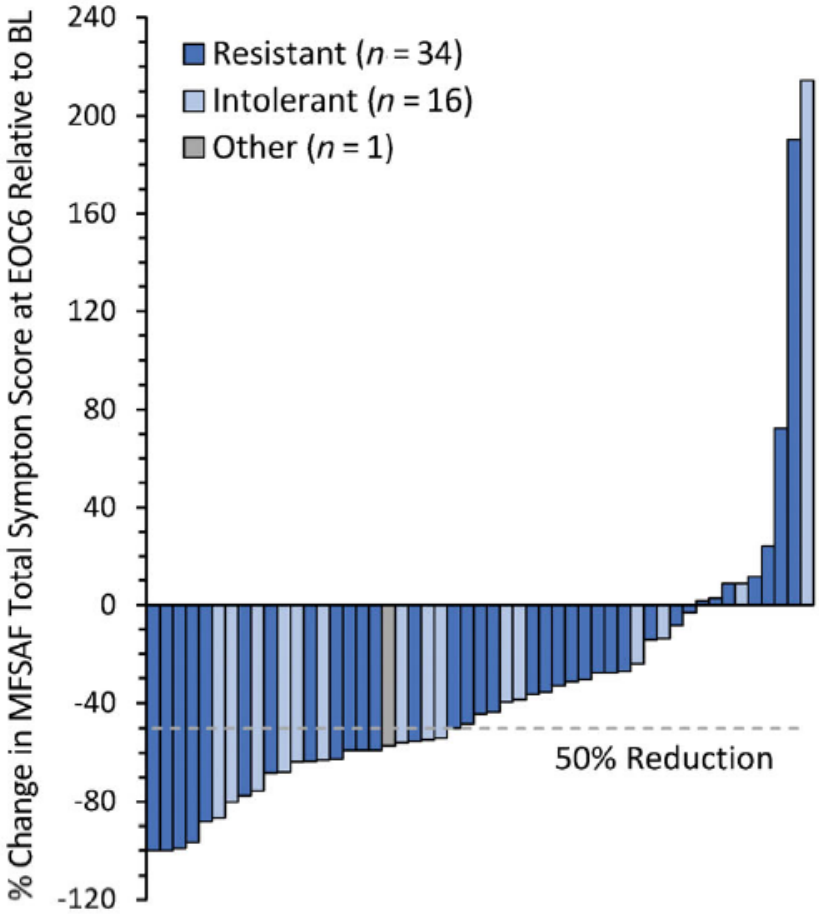


Figure 2: JAKARTA-2 ITT population TSS results (Harrison et al. 2020)[2]



The majority of non-responders in SIMPLIFY-2 as indicated by their TSS change from their baselines had a worsening of their symptoms. This would indicate that a positive utility increment (as suggested by the ERG in B2 and B3) would not be clinically plausible. A utility increment would in effect be suggesting that all patients who are R/R/I to ruxolitinib and continue their current therapy unchanged (i.e., continue BAT) enter the model and experience a benefit in QoL. This is unlikely and therefore the 0-utility applied for non-responders to BAT would be reasonable.

B3. Please add an option in the model so that it is possible to have no increment in utility for non-responders on BAT after fedratinib (as patients receive non-JAKi in the company base-case), but an increment in utility for patients initiated on BAT (as 88.5% are on ruxolitinib). This should look as below (using pre-EMA values).

Utility	Implementation	Female	Male	Source
Baseline		[redacted]	[redacted]	JAKARTA-2 (MF-8D)
JAKi response		[redacted]	[redacted]	JAKARTA-2 (MF-8D)
JAKi non-response		[redacted]	[redacted]	JAKARTA-2 (MF-8D)
BAT response (initiated on - 88.5% on rux)		[redacted]	[redacted]	JAKARTA-2 (MF-8D)
BAT non-response (initiated on - 88.5% on rux)		[redacted]	[redacted]	Assumption
BAT (non-JAKi) after Fed - response		0.000	0.000	
BAT (non-JAKi) after Fed - no response		0.000	0.000	
Worsening utility		-0.025	-0.025	Ruxolitinib SMC DAD

The company did not have sufficient time to make this model amendment. However, it is important to note that the company base case assumed the following:

- Patients initiated on fedratinib could not then respond to BAT (and therefore experience BAT non-responder utility)
- Non-responders to BAT experience zero utility increment

Therefore, using pre-EMA values, without any model amends, the above table would show the following output:

Utility	Female	Male	Source
Baseline	[redacted]	[redacted]	JAKARTA-2 (MF-8D)
JAKi response	[redacted]	[redacted]	JAKARTA-2 (MF-8D)
JAKi non-response	[redacted]	[redacted]	JAKARTA-2 (MF-8D)
BAT response (initiated on - 88.5% on rux)	[redacted]	[redacted]	JAKARTA-2 (MF-8D)
BAT non-response (initiated on - 88.5% on rux)	0.000	0.000	Assumption
BAT (non-JAKi) after Fed - response	0.000	0.000	Assumption
BAT (non-JAKi) after Fed - no response	0.000	0.000	Assumption
Worsening utility	-0.025	-0.025	Ruxolitinib SMC DAD

Further justification for the original assumptions are provided in the response to B2.

Section C: Textual clarification and additional points

EMA marketing authorisation

C1. The ERG understands that the EMA asked the company to provide an exploratory analysis re-classifying patients who were up-titrated (>400mg) as non-responders. The ERG read the EMA marketing authorisation and could not find wording suggesting that patients treated with fedratinib cannot have more than 400 mg. Please indicate the position and exact wording of the marketing authorisation that up-titration (dose > 400 mg) with fedratinib is not allowed.

The recommended dose of fedratinib is 400mg daily.[3]

The Summary of Product Characteristics (SmPC) does not explicitly state that 400mg daily is the maximum licensed dose, nor does it recommend dose-escalation for patients with an insufficient spleen and/or symptom response at the 400mg dose.

Dose modifications are only referred to in the SmPC in the context of managing treatment-emergent adverse reactions (haematologic toxicities, non-haematologic toxicities, and Wernicke's encephalopathy). Dose re-escalation is permitted for some toxicities once they have resolved up to the original dose level.

In the EMA CHMP report, it states for patients (n=33) in JAKARTA-2 study who were up-titrated to doses of 500mg and 600mg daily,[4] it is not clear if dose up-titration of fedratinib may have provided any additional clinical benefit for these patients.

It should also be noted that the spleen and symptom efficacy results presented in section 5.1 of the SmPC for the JAKARTA study (which compared fedratinib 400mg daily, fedratinib 500mg daily, and placebo) in JAK-inhibitor naïve patients relate only to the 400mg arm, despite including a 500mg arm.[3]

For FREEDOM 2 study which is currently ongoing in 5 UK centres, the daily dose of fedratinib cannot exceed 400 mg daily.[5]

For these reasons, whilst not explicit, it is implicit within the marketing authorisation that fedratinib 400mg daily is the maximum dose allowed. Anything above this dose is outside the marketing authorisation, therefore unlicensed and outside the scope of the appraisal.

References

1. Harrison CN, Vannucchi AM, Platzbecker U, et al. Momelotinib versus best available therapy in patients with myelofibrosis previously treated with ruxolitinib (SIMPLIFY 2): a randomised, open-label, phase 3 trial. *Lancet Haematol*. 2018; 5(2):e73-e81.
2. Harrison CN, Schaap N, Vannucchi AM, et al. Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the JAKARTA2 study using stringent criteria for ruxolitinib failure. *American Journal of Hematology*. 2020.
3. Celgene. Summary of Product Characteristics (SPC). Inrebic® (fedratinib) 100mg hard capsules. 2020. Available at: TBD. Accessed: TBD.
4. European Medicines Agency (EMA). Assessment report: Inebric. 2020. (Updated: 10 December 2020) Available at: https://www.ema.europa.eu/en/documents/assessment-report/inrebic-epar-public-assessment-report_en.pdf. Accessed: March 2021.
5. Clinicaltrials.gov. An Efficacy and Safety Study of Fedratinib Compared to Best Available Therapy in Subjects With DIPSS-intermediate or High-risk Primary Myelofibrosis, Post-polycythemia Vera Myelofibrosis, or Post-essential Thrombocythemia Myelofibrosis and Previously Treated With Ruxolitinib (FREEDOM2). 2020. (Updated: 10 March 2020) Available at: https://clinicaltrials.gov/ct2/show/NCT03952039. Accessed: 13 March 2020.

NICE National Institute for Health and Care Excellence

Patient organisation submission

Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis [ID1501]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type. [Please note that declarations of interests relevant to this topic are compulsory].

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	[redacted]

2. Name of organisation	Leukaemia Care
3. Job title or position	[redacted]
4a. Brief description of the organisation (including who funds it). How many members does it have?	Leukaemia Care is a national blood cancer charity, founded in 1969. We are dedicated to ensuring that anyone affected by blood cancer receives the right information, advice and support. Approximately 85-90% of our income comes from fundraising activities – such as legacies, community events, marathons etc. Leukaemia Care also received funding from a wide range of pharmaceutical companies, but in total those funds are less than 15% of our annual income. Leukaemia Care has undertaken a voluntary commitment to adhere to specific policies that regulate our involvement with the pharmaceutical industry set out at: http://www.leukaemiacare.org.uk/wp-content/uploads/2018/02/CODE-OF-PRACTICE.pdf
4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.]	2019/20: Celgene: £25,000 grant Bristol Myers-Squibb: £5,000 grant, £240 grant. Total = £5,240. Novartis: £25,000 grant, £447 grant, £11,792.95 grant, £7,279.69 grant. Total = £109,919.64

If so, please state the name of manufacturer, amount, and purpose of funding.	
4c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	N/A
5. How did you gather information about the experiences of patients and carers to include in your submission?	Information was gathered through the Leukaemia Care patient experience survey, which was last run in 2017. The survey included responses from 62 patients with myelofibrosis. We have also used the results from the International MPN Landmark Survey assessing the impact of MPN on patient quality of life and productivity: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5569657/. Further qualitative information and quotes also gathered from one to one discussion with myelofibrosis patients. Additionally, we have gathered information through our online forums, helpline, support groups and from communication with our membership. We also work closely with other patient groups and share expertise.
Living with the condition	
6. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	Myelofibrosis (MF) is a rare disorder of the bone marrow. It is one of the myeloproliferative neoplasms (MPN), a group of rare blood cancers. It is most common in patients aged over 50. **Diagnosis and emotional impact**

	A diagnosis of MF can have a huge impact on the patient's emotional well-being. In the 2017 Living with Leukaemia survey, many patients reported a change in their well-being; 51% of the patients felt depressed or anxious more often since their diagnosis. **Symptoms and impact on daily living** In the Leukaemia Care (2017) survey, the symptoms most commonly reported by MF patients since diagnosis includes fatigue (87%), fever/night sweats (65%), easily bruise or bleed (52%), feeling weak or breathless (50%), sleeping problem (50%), itchy skin (48%), pain in bones/joints (45%) and unexplained weight loss or loss of appetite (32%). The results from the international MPN landmark survey showed that 93% of MF patients with high symptom burden experienced a reduced quality of life, the highest percentage of the 3 MPNs studied. MF patients are likely to also have higher symptom burden compared to other MPN patients. These patients further reported that their MF caused emotional hardship (33%) and they felt worried or anxious about their disease (34%). Fatigue was also the most commonly reported symptom from the international MPN landmark survey. As a result, patients sometimes struggle to participate in daily life, such as exercising ***"Due to fatigue I cannot do anything physical or exercise, I also get breathless and end up coughing"***. The symptoms of MF, in particular fatigue, also have an impact on the patient's ability to work. 29% of the patients had to stop working and 21% had to reduce their working hours, according to the Leukaemia Care survey. Furthermore, 87% of the patients reported permanent long-term impact indicating that they are no longer able to work/continue education. Additionally, in the MPN landmark survey, many patients expressed that their disease had a high impact on daily activities and ability to work. - ***"My fatigue and anaemia had a lot of impact on my high intensity job as a doctor, I had to reduce hours"*** - ***"Fatigue greatly affected my quality of life, I had a managerial job and was quite drained at the end of the day, this continued after I was retired and throughout treatment with hydroxycarbamide and anagrelide"***

	Effect on carers According to the results from the MPN landmark survey, a higher number of MF patients reported to depend on a caregiver compared to patients with other MPNs. The study concluded this is likely due to high symptom burden observed in MF patients. Consequently, this increased dependence is likely to have an emotional impact on the caregiver/family member, as they will be required to take up extra responsibilities in order to support them. Additionally, some caregivers reported an impact on their employment due to reducing their hours in order to care for the individual with MF. The stress and physiological challenges associated with taking on these additional responsibilities can further have an impact on their relationship and their mental well-being.
Current treatment of the condition in the NHS	
7. What do patients or carers think of current treatments and care available on the NHS?	According to the Leukaemia Care survey, the physical effects most commonly reported by MF patients whilst on their most recent or current treatment include: fatigue (71%), itchy skin/rashes (33%), constipation or diarrhoea (31%), muscle, bone or joint pain (31%) and bleeding/bruising (25%). These are not too dissimilar to the symptoms reported at diagnosis, showing a need for effective treatments in this group. Additionally, 22% of patients reported that side effects had a large impact upon them and their life. When questioned about what they consider to be an important feature of a new treatment, 65% said tolerable side effects whilst on treatment and 79% said improved quality of life. 98% of MF patients surveyed would like a choice of different treatment options and 65% do not think there are enough treatment options currently available on the NHS. For fit patients, stem cell transplant is the only curative option. The only targeted therapy option for MF patients is ruxolitinib. Other treatments aim to control the symptoms patients experience as a result of their MF. In the front-line setting, treatments that are offered aim to control the MF symptoms. The impact these treatments have on patients varies. One patient in particular commented about the negative impact on quality of life due to treatment with hydroxycarbamide and anagrelide, which did not improve the

	symptoms of the disease, instead the patient found the treatments "***very toxic***"; "***I felt very unwell, no great appetite for life".*** This patient also experienced symptoms such as itchy skin and fatigue. Another patient commented on their treatment experience with interferon alpha, they felt the treatment ***"side effects was worse than any symptoms of MF"*** and that the treatment did not control any symptoms of their MF, instead they felt ***"constantly tired due to anaemia, spleen was getting bigger and losing weight"***. Fedratinib has shown to significantly reduce splenomegaly in clinical trials. This suggests the need for more targeted therapy options, as treatments aimed to specifically control the symptoms of MF may not be very effective for some patients and for these patients their quality of life is greatly affected. Patients experience of their recent or current treatment with ruxolitinib also varies, in one patient the symptoms of their MF were partially managed, and she described a good experience. However, one patient commented on side effects including weight gain, breathlessness, and infections. One patient on ruxolitinib continued to experience splenomegaly, which further resulted in requiring transfusions and greatly affected their ability to work. This highlights the way that patients differ in their response to particular treatments, emphasising the need for additional targeted treatment options to become available for patients with MF, to enable patients to make a choice. Additionally, further options are needed if and when these patients become resistant to ruxolitinib.
8. Is there an unmet need for patients with this condition?	There are very limited treatment options for MF patients that are unable to tolerate ruxolitinib or become intolerant/resistant overtime and are further unfit for stem cell transplant. Advances in research of targeted therapies means more treatment options should be made available for these patients, allowing access to alternative targeted treatments if they are unable to tolerate the current treatment options. This paper (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6935287/) further highlights the clear unmet need in patients that discontinue ruxolitinib, as they are likely to have "*dismal outcomes*".

Advantages of the technology	
9. What do patients or carers think are the advantages of the technology?	Fedratinib is a selective JAK2 inhibitor. In phase 3 JAKARTA and phase 2 JAKARTA-2 trials, Fedratinib has been shown to significantly reduce spleen volume and symptom burden in untreated patients and in patients previously treated with ruxolitinib respectively. As mentioned above, the symptoms experienced by MF patients and the side effects of their current treatments can have both a physical and emotional impact. In these patients, fedratinib offers a potential option in terms of controlling their MF symptoms and thus greatly impacting their quality of life. In the JAKARTA-2 trials, fedratinib was given to MF patients who were resistant or intolerant to prior ruxolitinib treatment. Keeping in mind the limited options available for patients that become resistant/intolerant to ruxolitinib, this new treatment provides an important alternative. Additionally, it is also shown to be effective in patients without prior ruxolitinib treatment, therefore likely provides an effective alternative in the front-line settings.
Disadvantages of the technology	
10. What do patients or carers think are the disadvantages of the technology?	Fedratinib has been linked to risk of serious encephalopathy, potentially due to thymine deficiency. However, this risk can be managed by monitoring thymine levels in patients prior to starting treatment and periodically during treatment. Other side effects of fedratinib correlate with ruxolitinib, including diarrhoea, nausea, anaemia and vomiting. The benefits and the need for this treatment in patients with very limited options outweighs the potential side effects.

Patient population	
11. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	All patients are likely to benefit from this treatment, due to very limited treatment options currently available for MF patients. This will further benefit patients that are unable to tolerate ruxolitinib and unfit for stem cell transplant.
Equality	
12. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	

Other issues	
13. Are there any other issues that you would like the committee to consider?	

Key messages

14. In up to 5 bullet points, please summarise the key messages of your submission:

- Myelofibrosis (MF) is a rare blood cancer with limited treatment options.
- The symptoms experienced by patients as a result of their MF can have a great impact on their quality of life, including on their ability to work in particular.
- Patients report that current non-targeted treatment options, such as hydroxycarbamide or interferon alpha, are not very effective and can instead result in worsening of their MF symptoms.
- There are very limited targeted treatment options available to MF patients. The only option is ruxolitinib, which can be effective but to which most patients are likely to become resistant or intolerant over time.
- In clinical trials, fedratinib has shown to be effective in terms of reducing splenomegaly and managing symptom burden in patients with and without prior ruxolitinib treatment. This will allow patients to benefit from an alternative effective option in both the front-line and relapsed/refractory settings, positively impacting their quality of life.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

NICE National Institute for Health and Care Excellence

Patient organisation submission

Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis [ID1501]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type. [Please note that declarations of interests relevant to this topic are compulsory].

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	[redacted]

2. Name of organisation	**MPN Voice**
3. Job title or position	[redacted]
4a. Brief description of the organisation (including who funds it). How many members does it have?	MPN Voice is the patient support organisation for people with Myeloproliferative Neoplasms (MPNs) in the UK. MPN Voice's mission is to provide clear and accurate information and emotional support to everyone who has been diagnosed with a myeloproliferative neoplasm and their families/friends. MPN Voice has members across the UK and in many other countries throughout the world. MPN Voice offers a website (http://www.mpnvoice.org.uk), patients' forums around the UK during the year, and a Peer Support programme to allow people with MPNs to contact others in similar circumstances. MPN Voice also has an online forum at HealthUnlocked which is a supportive and informative online forum where patients and carers can ask questions about anything related to MPNs, and get replies from people who really understand the challenges of living with a MPN. In addition, MPN Voice produces information leaflets and a newsletter for people with MPNs so that patients are better informed and have more confidence dealing with the management of their condition. MPN Voice also raises money to fund research towards a cure and advocacy for patients. MPN Voice's work is primarily funded by donations from the public, through a wide range of fundraising activities. MPN Voice also accepts financial support from pharmaceutical companies for specific activities (see below)

4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12	• MPN Voice has the following grants from **Novartis** over the past 12 months:
	  • 10/6/2019 £10,409.95 – support for Cork patient event
	  • 24/11/2019 £28,000 – support for Booklet printing and distribution and for patient events in 2019
	
	• We received the following grant from **Celgene** in 2019: £10,000 – support for National Patient Day

months? [Relevant manufacturers are listed in the appraisal matrix.] If so, please state the name of manufacturer, amount, and purpose of funding.	• An application to **Celgene** for financial support has been granted, but funds not yet received: £10,000 – support for patient events
4c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No
5. How did you gather information about the experiences of patients and carers to include in your submission?	Data supporting this submission has been gathered from a range of sources: MPN Voice is a founding member of MPN Advocates Network (MPNAN), a global coalition of MPN Patient groups. In 2019 MPNAN began the largest survey of MPN patient needs to date, with over 1700 responses at the time of writing. 302 responses have been received from myelofibrosis patients. Evidence has also been taken from two MPN Landmark studies, the original US-based one in 2016 and a subsequent international study. The 2016 study had 816 respondents, of which 2017 were Myelofibrosis patients. The international study had 174 responses from myelofibrosis patients, 45 from the UK, and provides information on patient reported quality of life and productivity. (Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5569657/) This submission is also informed by a patient experience survey of 34 adults diagnosed with myelofibrosis, carried out by Leukaemia Care in 2016. This was part of a wider survey of over 2500 blood cancer patients.

	MPN Voice continually gathers information through our support services (helpline, support groups, conferences, communications with our membership) and one to one discussion with patients.
Living with the condition	
6. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	Myelofibrosis (MF) is a rare form of blood cancer, known as a myeloproliferative neoplasm (MPN), that causes the overproduction of fibroblasts in the bone marrow. There are fewer than 1-2 people per 100,000 diagnosed every year in the UK. Most patients will be over the age of 50 years old at diagnosis, with the average age in the Landmark study being 59.6 years old. There are two types of myelofibrosis, primary and secondary. In primary MF the disorder has arisen by itself and secondary MF is a progression from another MPN. Around 50-60% of MF patients will have a mutation in the JAK2 protein. The international MPN Landmark study performed a systematic analysis of the burden of MPN illnesses. Quoting from the peer-reviewed report of the study, "*MPNs are associated with a substantial disease burden, often leading to a reduced quality of life (QOL) for many patients. Symptoms may include fatigue, pruritus, night sweats, microvascular symptoms, splenomegaly, and splenomegaly associated symptoms (e.g., abdominal pain, early satiety), with fatigue being one of the most severe symptoms. Among patients with MF, PV, or ET, patients with MF generally have the highest symptom burden and the lowest QOL*." MF patients reported to the 2016 Landmark researchers a range of symptoms. The following are illustrations of the numbers of patients for whom the symptoms have a significant impact: • Fatigue 80% of patients • Depression or sad mood 75% • Abdominal discomfort 53% • Night sweats 51%

	Apart from the actual symptoms, MF affects many other aspects of patients' lives. The MF patients in the UK who responded to the MPNAN survey scored 4.2/10 in terms of financial impact (0 being the most significant impact). Over 30% of these patients reported significant financial difficulties. The impact of the disease is also felt by the people who care for MF patients. This impact is felt in a variety of ways, from the psychological and emotional burden of caring for someone with an incurable, debilitating disease, to the practical and financial effect. On average respondents to the MPNAN survey who specifically identified as carers of MF patients scored 6.7/10 for the impact on their ability to work (10 meaning they couldn't work at all), and over 30% reported that they were unable to work at all because of their role as carers. The disease significantly impacts the economic productivity of patients and their carers. The 2016 Landmark survey reported that 59% of MF patients had reduced work hours owing to the disease.
Current treatment of the condition in the NHS	
7. What do patients or carers think of current treatments and care available on the NHS?	Following diagnosis, patients who aren't experiencing symptoms will be put on 'Watch and Wait' where the MF is monitored over time. In the Leukaemia Care (LC) survey, 29% of patients were placed on Watch and Wait and this caused some level of concern or worry for the majority (62%) of patients. Overall, 62% of MF patients felt to some extent more depressed or anxious following diagnosis, including those who had started treatment or were still on Watch and Wait, demonstrating the significant emotional impact that a diagnosis has on the patient. Other MF patients will be given treatments to manage MF and the side effects, as the only curative option is stem cell transplant. With this being an intensive treatment option, it is not often advised. Just 9% of patients in the Leukaemia Care survey had received a stem cell transplant. LC asked about the side effects of their current treatments, the majority of patients experienced side effects (94%) with the most common being: fatigue (68%), sleeping problems (41%), bruising (41%), sore mouth (38%), anaemia (35%), loss of concentration/memory (32%), and breathing difficulties (32%). The side effects had an impact on 82% of patients (54% small impact, 25% large impact, 4% intolerable).

	LC also gained anonymous evidence from three patients about their treatment with ruxolitinib. The degree to which the treatment impacted on their symptoms was very different, with one patient saying symptoms had gotten worse, and the others stating symptoms had partially or significantly improved. One patient stated that they failed to respond to ruxolitinib after 2-3 years and their spleen enlarged. This was their most recent treatment for MF, demonstrating the lack of options for patients.
8. Is there an unmet need for patients with this condition?	Most therapies for MF focus on controlling the symptoms of the disease and these therapies are not effective for all MF patients; many patients do not tolerate their side effects well. Ruxolitinib treatment is effective for some patients, but response if frequently inadequate. Furthermore, the median duration of response to ruxolitinib is 3 years and we are seeing increasing numbers of patients with progressive disease after previous response to ruxolitinib. To quote from the Dec 2019 paper *Beyond Ruxolitinib: Fedratinib and Other Emergent Treatment Options for Myelofibrosis*, "…patients who discontinue ruxolitinib have dismal outcomes, making this situation an area of significant unmet need" This patient group (those who need to discontinue ruxolitinib treatment) represents an area of major unmet medical need as currently there are no approved therapies for this patient group in the UK.
Advantages of the technology	
9. What do patients or carers think are the advantages of the technology?	As described above, MF has a serious impact on MF patients' quality of life and treatments are limited. Ruxolitinib was approved as a therapy in 2012 but there have been no new therapies since that time. If patients either do not tolerate ruxolitinib, or the effectiveness of ruxolitinib declines over time, fedratinib can be an effective option. Currently, fedratinib presents the only approved therapy for patients who have experienced a treatment failure with ruxolitinib. Fedratinib has been shown to reduce spleen size in over half of patients with an inadequate response to ruxolitinib with at least a 50% reduction in spleen length in 34%. Fedratinib also improved disease related symptoms in this patient group. Trials have also shown that fedratinib is effective in frontline use, with approximately one third of patients experiencing a substantial reduction in spleen size after 24 weeks of

	therapy, with an improvement in symptoms in a similar proportion.
Disadvantages of the technology	
10. What do patients or carers think are the disadvantages of the technology?	Fedratinib has side effects, including diarrhoea, nausea, anaemia and vomiting. These have been modified with extra medications. There have been instances of patients developing Wernicke encephalopathy, whilst taking fedratinib on clinical trials. This is a potentially life-threatening complication, but this complication is rare and the direct relationship to fedratinib remains uncertain in this patient group who are predisposed to nutritional deficiencies that increase the risk of Wernicke encephalopathy irrespective of fedratinib treatment. Risk of this complication can be mitigated by careful selection and close monitoring of patients who are prescribed fedratinib. Overall, Fedratinib has a favourable risk/benefit ratio.
Patient population	
11. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	As described above, fedratinib is a potentially useful option (where very few others exist) in situations where other treatments have failed or cannot be tolerated by patients.

Equality	
12. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	No
Other issues	
13. Are there any other issues that you would like the committee to consider?	No
Key messages	
14. In up to 5 bullet points, please summarise the key messages of your submission: • Myelofibrosis is a debilitating disease that has a significant impact on patients' quality of life • The impact of the disease is felt by patients' carers as well as by the patients themselves and has significant social and economic effects • The only cure for MF is a stem cell transplant, which is not an option for most patients. Many patients do not tolerate existing therapies and therefore need other options	

- The only other targeted therapy for MF is ruxolitinib which, if it works, is only effective for a few years. Patients need an option for subsequent treatment
- Fedratinib has been shown to improve the quality of MF patents' lives and should be available as an option

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

Fedratinib for splenomegaly and symptoms in myelofibrosis. A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK Rachid Rafia, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK Andrew Rawdin, Health Economic Modeller, ScHARR, University of Sheffield, Sheffield, UK Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK
Correspondence Author	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
Date completed	15/09/2020

Source of funding: This report was commissioned by the NIHR HTA Programme as project number GID-TA10406.

Declared competing interests of the authors

None of the authors has any conflicts of interest to declare.

Acknowledgements

We would like to thank Sunil Gupta (Lewisham and Greenwich NHS Trust), Dragana Milojkovic (Imperial College Healthcare NHS Trust), and Joel Newman (East Sussex Healthcare NHS Trust), for clinical advice relating to this project. We would also like to thank Hazel Squires, ScHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Martyn-St James M, Rafia R, Stevens J, Rawdin A, Wong J. Fedratinib for splenomegaly and symptoms in myelofibrosis: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2020.

Contributions of authors

Rachid Rafia and Andrew Rawdin critiqued the health economic analysis submitted by the company. Marrissa Martyn-St James summarised and critiqued the clinical effectiveness data reported within the company's submission. John Stevens critiqued the statistical aspects of the submission. Ruth Wong critiqued the company's search strategy. All authors were involved in drafting and commenting on the final report.

CONTENTS

List of tables

List of figures

List of Boxes

Abbreviations

AE	Adverse event
AiC	Academic in confidence
ASCT	Allogeneic Stem Cell Transplant
BAT	Best Available Therapy
BIC	Bayesian Information Criterion
BNF	British National Formulary
BSA	Body Surface Area
CI	Confidence interval
CS	Company submission
CT	Computed tomography
DES	Discrete event simulation
DIPSS	Dynamic International Prognostic Scoring System
DSA	Deterministic sensitivity analysis
DSU	Decision Support Unit
ECOG PS	Eastern Cooperative Oncology Group performance status
EMA	European Medicines Agency
eMIT	electronic Market Information Tool
EOC3	End of cycle three
EOC6	End of cycle six
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer quality of life questionnaire
ERG	Evidence Review Group
FAD	Final Appraisal Document
FDA	Food and Drug Administration
GRA	Global response assessment
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
ICER	Incremental cost effectiveness ratio
IPSS	International Prognostic Scoring System
ITC	Indirect treatment comparison
ITT	Intention-to-treat
JAK	Janus kinase inhibitor
KM	Kaplan-Meier
L	Litre
LOCF	Last observation carried forward

MAIC	Matching-Adjusted Indirect Comparison
MF-SAF	MF-SAF, myelofibrosis symptom assessment form
MIMS	Monthly Index of Medical Specialities
mg	Milligram
mL	Millilitre
mm	Millimetre
MRI	Magnetic resonance imaging
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
ONS	Office for National Statistics
OS	Overall survival
PAS	Patient Access Scheme
PSA	Probabilistic sensitivity analysis
PP	Per-protocol
PSS	personal and social services
QALY	Quality-adjusted life-year
RBC	Red blood cell
RCT	Randomised controlled trial
RR	Relative risk
SCT	Stem cell transplant
SD	Standard deviation
SLR	Systematic literature review
SmPC	Summary of Product Characteristics
STA	Single Technology Appraisal
STC	Simulated Treatment Comparison
SVR	Spleen volume response
TA	Technology Assessment
TEAE	Treatment-Emergent Adverse Event
TSS	Total symptom score
TTD	Time to Treatment Discontinuation
VB	Visual Basic
WHO	World Health Organisation

1 EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the evidence review group (ERG) as being potentially important for decision making. It also includes the ERG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, technology and evidence and information on non-key issues are in the main ERG report.

All issues identified represent the ERG's view, not the opinion of NICE.

1.1 Overview of the ERG's key issues

Table 1: Overview of the ERG's key issues

ID	Summary of issue	Report sections
Issue 1	Concerns with the Phase 2, single-arm, JAKARTA-2 study: estimates of treatment effects from Phase 2 studies such as JAKARTA-2 generally over-estimate true treatment effects, populations defined by Phase 2 studies and tend to generate larger treatment effects, lack of a concurrent control in JAKARTA-2 means that the study is likely to suffer from the phenomenon known as regression to the mean, it is not clear why JAKARTA-2 did not include a concurrent control, although the primary outcome measure may be considered clinically relevant, dichotomising a continuous variable is statistically inefficient	3.2.5
Issue 2	Estimates of relative treatment effect may be biased and over-precise. Estimates of relative treatment effect are presented on the absolute risk scale	3.4.3
Issue 3	Costs and efficacy inputs are not aligned in the economic model – ICER should be presented for each population separately	4.3.4.1
Issue 4	Inappropriate modelling approach led to inaccurate estimation for TTD and prediction for the time in health states	4.3.4.2
Issue 5	Ommission of supportive care and inappropriate assumptions for HRQoL in patients initiated on fedratinib	4.3.4.3.1 and 4.3.4.13
Issue 6	Inconsistent assumptions between fedratinib and comparator for the duration on treatment	4.3.4.4
Issue 7	Absence of robust evidence to support a survival difference between fedratinib and BAT	4.3.4.5and 4.3.4.6
Issue 8	Lack of face validity for the scenario presented using the stopping rule; which reduce costs but not OS	4.3.4.10
Issue 9	Uncertainty regarding the costing for ruxolitinib	4.3.4.11
Issue 10	Lack of reliability of response rates used in the economic model	4.3.4.7
Issue 11	End of life criteria not met in the company's base-case	6

The ERG's critical appraisal identified several important issues relating to the company's model and the evidence used to inform its parameters. While some of these issues are debatable and may reflect matters of subjective opinion, others reflect more serious underlying problems regarding the conceptualisation and implementation of the model and the use of evidence to inform the model's parameters. Consequently, the ERG has serious doubts regarding the validity of the results presented within the CS and would advise considerable caution in their interpretation and use in informing decision-making. These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethinking of the

model's logic. As such, the impact of these issues on the expected cost-effectiveness of fedratinib is not clear.

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- assuming that fedratinib is associated with a survival difference compared with best available treatments (BAT), assumed to be comprised mostly of ruxolitinib in the economic model (89%),
- improving quality of life (improvement in symptoms) through higher response rates (as estimated by the company).

Overall, the technology is modelled to affect costs by:

- Higher unit price than current treatments,
- Reduction in resource use,
- Assuming different duration on treatment for patients initiated on ruxolitinib/BAT (comparator) or fedratinib.

The modelling assumptions that have the greatest effect on the ICER are:

- The comparator and population entering the economic model assumed; in particular the BAT composition (the proportion of patients assumed to receive ruxolitinib),
- How long patients on BAT (comprising mostly of ruxolitinib) and fedratinib are assumed to remain on treatment,
- The size of overall survival benefit, if any
- Assumptions regarding HRQoL following JAK inhibitors discontinuation (in patients no longer treated with JAK inhibitors),
- Assumptions regarding costing for the comparator arm (dosage received for patients treated with ruxolitinib).

1.3 The decision problem: summary of the ERG's key issues

The Evidence Review Group (ERG) considers the company's description of the underlying health problem in the company's submission (CS) to be appropriate, up-to-date and mostly relevant to the decision problem set out in the final National Institute for Health and Care Excellence (NICE) scope. The target population in the CS is people with primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis that have been treated with Ruxolitinib, which is a subset of the population in the NICE scope (people with primary myelofibrosis, post polycythaemia vera myelofibrosis or post essential thrombocythaemia myelofibrosis). The

comparator in the CS is established clinical practice, otherwise referred to as BAT (including but not limited to Ruxolitinib, hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin, and RBC transfusion). The ERG has no key issues with the decision problem.

1.4 The clinical effectiveness evidence: summary of the ERG's key issues

The ERG considers the searches for clinical effectiveness evidence reported in the CS to be adequate, and believes the Phase 2, included single-arm study of fedratinib (JAKARTA-2) to be relevant to the NICE decision problem. However, the ERG has some concerns with the design of the JAKARTA-2 study and using evidence from it to make reimbursement decisions in the target population.

The ERG also has some concerns with the unanchored indirect comparison of fedratinib to BAT (best available therapy), using two RTCs of other JAK2 inhibitors (the PERSIST-2 RCT and SIMPLIFY-2 RCT).

Issue 1. Concerns with the Phase 2, single-arm, JAKARTA-2 study

Report section	Section 3.2.5
Description of issue and why the ERG has identified it as important	1. Only those Phase 2 studies with good results lead to treatments being evaluated in Phase 3 studies. Consequently, estimates of treatment effects from Phase 2 studies such as JAKARTA-2 generally over-estimate true treatment effects. 2. Populations defined by Phase 2 inclusion/exclusion criteria are often less diverse than in Phase 3 studies and tend to generate larger treatment effects. 3. The lack of a concurrent control in JAKARTA-2 means that the study is likely to suffer from the phenomenon known as regression to the mean such that recruitment to the study is a consequence of extreme values that return to their average values post-treatment even if there is no treatment effect. 4. It is not clear why JAKARTA-2 did not include a concurrent control. The established clinical practice for patients treated with ruxolitinib in the UK includes a basket of treatment options that are supportive but do not alter the course of disease. 5. Although the primary outcome measure may be considered clinically relevant, dichotomising a continuous variable is statistically inefficient.
What alternative approach has the ERG suggested?	Some of the limitations associated with the design of the study can be mitigated by adjusting for all relevant prognostic factors and treatment effect modifiers. However, as discussed in Section 3.4 of this ERG report, it is impossible to specify the correct model and there will be residual bias.
What is the expected effect on the cost-effectiveness estimates?	This is unclear

What additional evidence or analyses might help to resolve this key issue?	Ongoing studies identified by the company include the Phase III, single-arm FREEDOM study of fedratinib in patients with intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib is currently recruiting and due to read out in 2022 (US study); and the Phase III FREEDOM-2 study of fedratinib compared with BAT in patients with intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib is currently recruiting and due to read out in 2022. These studies can potentially mitigate the above limitations associated with the design of JAKARTA-2

Issue 2. Concerns with the unanchored indirect comparison of fedratinib to BAT

Report section	Section 3.4
Description of issue and why the ERG has identified it as important	The ERG believes that it is unlikely that the propensity score type model has accounted for all prognostic factors and treatment effect modifiers and that estimates of relative treatment effect may be biased and over-precise. 1 It is unclear whether there are differences in patient populations beyond the attempt to adjust for all relevant prognostic factors and treatment effect modifiers. 2 There are concerns with some aspects of the methods used to identify prognostic factors and treatment effect modifiers Estimates of relative treatment effect are presented on the absolute risk scale rather than the odds ratio (or log-odds ratio scale) which is assumed to be the additive scale on which relative treatment effects are estimated.
What alternative approach has the ERG suggested?	The ERG suggests that variables included in the International Prognostic Scoring System (IPSS) and the Dynamic International Prognostic Scoring System (DIPSS) should be included in the model irrespective of their statistical significance. Continuous variables should be included as continuous variable if possible. The absolute response to fedratinib should be computed by adding the relative effect to the BAT response on the additive scale and not by subtracting the difference in absolute responses from the fedratininb response as the company has done.
What is the expected effect on the cost-effectiveness estimates?	The impact is unclear
What additional evidence or analyses might help to resolve this key issue?	As stated above, the ERG suggests that variables included in the IPSS and the DIPSS should be included in the model irrespective of their statistical significance. Continuous variables should be included as continuous variable if possible. The absolute response to fedratinib should be computed by adding the relative effect to the BAT response on the additive scale and not by subtracting the difference in absolute responses from the fedratininb response as the company has done.

1.5 The cost-effectiveness evidence: summary of the ERG's key issues

Issue 3. Alignement between the comparator assumed and the population entering the economic model

Report section	Section 4.3.4.1
Description of issue and why the ERG has identified it as important	The population entering the economic model appear to be a mix of (1) patients who switch to fedratinib at the point where they become resistant/intolerant to ruxolitinib but are continued on ruxolitinib and (2) patients who switch to fedratinib at the point when they discontinued ruxolitinib due to AEs or lack of efficacy. BAT is a composite comparator and consist of mostly ruxoltinib (89%), therefore implying that most patients entering the model relate to population 1 and therefore switch to fedratinib at the point of resistance/intolerance, not at the point of discontinuation. Key efficacy input (OS) is not aligned with the population entering the model and is taken from people no longer treated with ruxolitinib at the point when they discontinue.
What alternative approach has the ERG suggested?	To allow better alignment between costs and other inputs in the economic model but also facilitate interpretation of results, the ERG consider that two analysis should be presented for the two separate populations, as the comparator would be different. - For patients who switch to fedratinib at the point where they become resistant/intolerant to ruxolitinib, the comparator is ruxolitinib alone (or in combination) - In patients who discontinue ruxolitinib, the comparator is BAT without ruxolitinib
What is the expected effect on the cost-effectiveness estimates?	This is unclear
What additional evidence or analyses might help to resolve this key issue?	ICERs needs to presented for two population separately. In particular, given that 89% of patients are assumed to receive ruxolitinib (mostly costly and effective compared with other BAT treatments), it is relevant to assess fedratinib versus ruxolitinib alone and ensure that costs are aligned with efficacy inputs.

Issue 4. Inappropriate approach to modelling

Report section	Section 4.3.4.2
Description of issue and why the ERG has identified it as important	The approach to modelling chosen by the company (individual-event based model, with patients separated onto group categories) lead to a number of biases: - Inaccurate estimation of TTD for fedratinib (Section 4.3.4.2.2) and introduction of unnecessary uncertainties (4.3.4.2.3), - Inaccurate estimation of time in health state (Section 4.3.4.2.4) which does not align with the trial data, - inaccurate prediction for the AML health state (Section 4.3.4.2.5 and Section 4.3.4.3.2)
What alternative approach has the ERG suggested?	In addition to concerns regarding the structure, the ERG has a number of other concerns regarding the company's implemented economic model. While some of these issues are debatable and may reflect matters of subjective opinion, others reflect more serious underlying problems regarding the conceptualisation and implementation of the model and the use of evidence to inform the model's parameters. The ERG therefore considers that the joint resolution of these problems would require a 'full' rethinking of the model's logic.
What is the expected effect on the cost-effectiveness estimates?	These structural problems are intertwined with other conceptual issues, and the resolution of this issue in isolation would not result in an appropriate or credible model. The impact on the expected cost-effectiveness of fedratinib is therefore not entirely clear. To illustrate those, for instance, using some of the ERG preferred assumptions, when assuming no survival difference, the company estimate that fedratinib is dominant (e.g., is associated with less costs and is more effective). However, when TTD is corrected the ICER change from being dominant to £444,999 per QALY gained. There are also broader conceptual issues (how long patients remain on treatment and health states) which need to be addressed at the same time.
What additional evidence or analyses might help to resolve this key issue?	The ERG considers that the joint resolution of problems identified would require a 'full' rethinking of the model's logic

Issue 5. Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib

Report section	Section 4.3.4.3.1 and Section 4.3.4.13
Description of issue and why the ERG has identified it as important	No supportive care health state is considered in the model. Patients are assumed to remain on BAT for life which the ERG does not consider appropriate in line with TA386 and data from the HMRN. The company also assumes that after fedratinib discontinuation, patients receive BAT (consisting mostly of HU – non-JAKs) but experience HRQoL similar to patient treated with JAK (non-responders and responders to JAKs are assumed to be the same as non-JAKs) It is also more likely that after 2 JAKs patients may enter supportive care health state, where HRQoL will worsen over time.
What alternative approach has the ERG suggested?	Due to limitations with the company's model structure, it was not possible for the ERG to include supportive care and explore this.
What is the expected effect on the cost-effectiveness estimates?	The current assumption in the model is likely to be favourable to to fedratinib, as patients are assumed to discontinue JAK and move onto non-JAK treatments maintaining improvement in HRQoL (but low costs). In the comparator arm, patients are on JAK and experience same quality of life, but high costs.
What additional evidence or analyses might help to resolve this key issue?	The ERG considers that the joint resolution of problems identified would require a 'full' rethinking of the model's logic

Issue 6. Inconsistent assumption between BAT and fedratinib

Report section	Section 4.3.4.4
Description of issue and why the ERG has identified it as important	The company assumes that patients initiated on BAT (comparator arm, comprising of 89% ruxolitinib) remain in this health state for life (unless AML or last 8 weeks of life), incurring high costs. This is justified by the company by the absence of alternative targeted therapy following ruxolitinib discontinuation. In contrast, patients initiated on fedratinib discontinue early as per trial (using TTD), and move onto BAT (mostly HU) spending a similar amount of time but incurring low costs. Inconsistent assumptions are used between treatment arms and this is inconsistent with the key CS argument that patients are kept on ruxolitinib while achieving a suboptimal response in the absence of alternative therapy.
What alternative approach has the ERG suggested?	Assumptions need to be consistent between treatment arms. As per assumption made for BAT/ruxolitinib (and argument advanced by the CS), the ERG considers that patients on fedratinib should continue treatment (fedratinib sub-optimally) beyond TTD A full' rethinking of the model's logic is required
What is the expected effect on the cost-effectiveness estimates?	Using the company's base-case, the ICER increases from £11,645 (10,000 patients) to £168,781 (assuming all patients [responders and non-responders] remain on fedratinib for life – as assumed for the comparator arm). Analysis is limited by the current model structure
What additional evidence or analyses might help to resolve this key issue?	A 'full' rethinking of the model's logic is required to ensure the model is conceptually valid

Issue 7. Assumption of survival difference

Report section	Section 4.3.4.5and 4.3.4.6
Description of issue and why the ERG has identified it as important	OS for fedratinib and comparator is taken from two separate sources and is not adjusted for differences in patient characteristics, so the final comparison is a naïve indirect comparison. Key evidence used by the company (taken from patients no longer treated with ruxolitinib at the point of discontinuation) also does not appear to relate to the population entering the economic model (mostly patients continuing ruxolitinib at the point when they "should" discontinue ruxolitinib). It is the ERG's view that evidence from SIMPLIFY-2, COMFORT-trials and JAKARTA are not suggestive of a difference in survival between fedratinib and BAT
What alternative approach has the ERG suggested?	In light of evidence presented by the company and identified by the ERG, it is the ERG's view that no survival difference should be assumed.
What is the expected effect on the cost-effectiveness estimates?	The ICER improve using the company's base-case assumption when assuming no difference in survival (because of inconsistent assumptions between arms, but also errors in the economic model). The ERG analysis show that the ICER assuming no survival difference is improved prior correction of error for TTD. When the error is corrected, the ICER change from being dominant to £444,999 per QALY gained.
What additional evidence or analyses might help to resolve this key issue?	Comparing OS from SIMPLIFY-2 (KM value at 24 weeks) versus JAKARTA-2 in a similar population (adjusted for baseline characteristics). Provide comparisons to the COMFORT-trials for ruxolitinib (adjusted for differences).

Issue 8. Lack of face validity for the scenario presented using the stopping rule

Report section	Section 4.3.4.10
Description of issue and why the ERG has identified it as important	It is not entirely clear if a stopping rule would apply in UK clinical practice for fedratinib given that patients already failed a JAK (ruxolitinib) and there are no alternative therapeutic options available if fedratinib is available. Assuming a stopping rule is also inconsistent with the key CS assumption that patients continue ruxolitinib sub-optimally. In the economic model, OS is unchanged using the stopping rule. This is unsupported. This is also conceptually inconsistent with the company's model prediction. This is also not inconsistent with the approach in TA386.
What alternative approach has the ERG suggested?	If a stopping rule is considered relevant, the impact of such stopping rule on OS need to be reflected as in TA386.
What is the expected effect on the cost-effectiveness estimates?	ICERs will deteriorate compared with those presented by the company for this scenario (as costs as cut, but OS benefit are maintained)
What additional evidence or analyses might help to resolve this key issue?	A new analysis is required should the stopping rule scenario be considered for decision-making.

Issue 9. Costings for the comparator arm (ruxolitinib)

Report section	Section 4.3.4.11
Description of issue and why the ERG has identified it as important	Ruxolitinib is costed based on the platelet count distribution in JAKARTA-2. Platelet count distribution in SIMPLIFY-2 (used for response rate for BAT) is likely (not reported) to be lower (given differences in mean platelet counts). Platelet count distribution in other studies identified as part of the CS systematic review report less favorable platelet count distribution (for costing).
What alternative approach has the ERG suggested?	It is unclear to the ERG if the platelet count distribution in JAKARTA-2 is reflective of patients that would receive fedratinib in UK practice, and what is the platelet count distribution in SIMPLIFY-2 (which is used for response rate for BAT).
What is the expected effect on the cost-effectiveness estimates?	The ICER is likely to deteriorate if the platelet count distribution from SIMPLIFY-2 was used.
What additional evidence or analyses might help to resolve this key issue?	Platelet count distribution needs to match evidence for efficacy.

1.6 Other key issues: summary of the ERG's view

Issue 10. Reliability of response rate

Report section	Section 4.3.4.7
Description of issue and why the ERG has identified it as important	• Concern with the comparability of the population included in SIMPLIFY-2 and JAKARTA-211 despite an attempt from the company to adjust for baseline characteristics, • Concern with comparability in design which is likely to be more favourable to fedratinib (no washout period in SIMPLIFY-2) • Concern with estimation of the treatment effect using difference in absolute risk and its application to baseline risk on the absolute risk scale. • Assumption that response rate for BAT after fedratinib is the same as BAT (inclusion fedratinib) • Use of spleen response by volume less favorable to the comparator arm (no difference for fedratinib)
What alternative approach has the ERG suggested?	In addition to adjustments for baseline characteristics, the population for fedratinib needs to reflect the population included in SIMPLIFY-2 (likely to be mostly introlerant or relapse). Exploration of the impact of using spleen length (by palpitation) instead of volume.
What is the expected effect on the cost-effectiveness estimates?	The impact is unclear, but current assumptions are more likely to be de-favourable to the comparator arm, and therefore, the ICER is likely to deteriorate.
What additional evidence or analyses might help to resolve this key issue?	• As stated above, in addition to adjustment for baseline characteristics, the population for fedratinib needs to reflect the population included in SIMPLIFY-2 (likely to be mostly intolerant or relapsed). • Exploration of the impact of using spleen length (by palpitation) instead of volume.

Issue 11. End of life criteria

Report section	Section 6
Description of issue and why the ERG has identified it as important	The company's base-case predicts a mean LY over 2 years (years), and therefore does not meet EoL criteria. As described, the ERG does not consider evidence available to be suggestive of a survival difference between fedratinib and BAT and therefore the ERG is unclear if fedratinib would be associated with a gain in survival of 3 month.
What alternative approach has the ERG suggested?	The mean LY for BAT using the ERG prefered assumption is considerably higher to those estimated by the company and therefore over 2 years, and therefore does not meet EoL criteria No survival difference is assumed by the ERG in light of the evidence available, and therefore does not meet EoL criteria
What is the expected effect on the cost-effectiveness estimates?	NA
What additional evidence or analyses might help to resolve this key issue?	While uncertain, the EoL criteria is not met in the company base-case model (mean LY of years)

1.7 Summary of ERG's preferred assumptions and resulting ICER

Table 2. Summary of ERG preferred assumptions and ICER

Scenario	Incremental cost	Incremental QALYs	ICER (change from company base case)
Company's base case (assuming 10,000 patients)	[redacted]	[redacted]	£11,645
Analysis 1: Disabling of options	[redacted]	[redacted]	£8,477
Analysis 1 + Analysis 2: Change to parametric distributions	[redacted]	[redacted]	£8,303
Analysis 1-2 + Analysis 3: assumption of no survival difference between fedratinib and BAT	[redacted]	[redacted]	Dominant
Analysis 1-3 + Analysis 4: Correction of errors for TTD due to the inappropriate modelling approach	[redacted]	[redacted]	£444,999
Analysis 1-4 + Analysis 5: Assumption that 88.5% of responders on fedratinib remain on treatment (sub-optimally) for life	[redacted]	[redacted]	£1,382,748
Analysis 1-5 + Analysis 6: Assumption that 88.5% of both responders and non-responders on fedratinib remain on treatment (sub-optimally) for life [as per assumption for comparator arm – where patients are assume to remain on suboptimal ruxolitinib for life]	[redacted]	[redacted]	£2,959,869
ERG preferred base-case (Analysis 1 – 6)	[redacted]	[redacted]	**£2,959,869**

Modelling errors identified and corrected by the ERG are described in Section 4.3.3.1. For further details of the exploratory and sensitivity analyses done by the ERG, see Section 4.4 and Section 5.

Owing to the ERG's concerns regarding the robustness of the company's model, the results generated using the company's model, including the ERG's exploratory analyses, should be interpreted with caution. However, the exploratory analysis conducted by the ERG illustrates the likely impact on the ICER and the uncertainty around the analysis and its results.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The Evidence Review Group (ERG) considers the company's description of the underlying health problem in the company's submission (CS)[1] to be appropriate, up-to-date and mostly relevant to the decision problem set out in the final National Institute for Health and Care Excellence (NICE) scope.[2] The ERG provides a brief summary of the underlying health problem in this section.

Clinical features

The clinical features of myelofibrosis include progressive anaemia, leucopenia or leucocytosis, thrombocytopenia or thrombocytosis and multi-organ extramedullary haemopoiesis; which commonly cause hepatomegaly and symptomatic splenomegaly.[3] Progressive bone marrow fibrosis results in release of the malignant stem cells into the circulation that can result in extramedullary haematopoiesis manifesting as splenomegaly.[3]

The disease is characterized by a clonal haemopoietic stem cell proliferation associated with a characteristic stromal pattern, a leuco-erythroblastic blood film and elevated levels of various inflammatory and pro-angiogenic cytokines.[3]

Patients with advanced disease experience severe constitutional symptoms, the consequences of massive splenomegaly (pain, early satiety, splenic infarction, portal hypertension and dyspnoea), progressive marrow failure, pulmonary hypertension, transformation to leukaemia and early death.[3]

Allogeneic stem cell transplant is the only potentially curative treatment for myelofibrosis, however, it is only suitable for people who are fit enough to undergo treatment.[2]

Aetiology

Citing work by Romano et al. (2017)[4] and Song et al. (2018),[5] the CS (Page 12) reports that "*most patients with myelofibrosis have a mutation that results in constitutive activation of the JAK/Signal Transducer and Activator of Transcription (STAT) signalling pathway*".[1]

Myelofibrosis can present as a *de novo* disorder (primary MF) or evolve secondarily from previous polycythaemia vera or essential thrombocythaemia (Post-PV MF or Post-ET MF respectively).[3]

Prevalence

MF prevalence estimates in the UK for 1995-2012 are reported as being 0.9 cases per 100,000.[6] Around 2 to 3 people per 100,000 are diagnosed with myelofibrosis every year.[7] The median survival is 5 years from onset, but variation is wide; some patients have a rapidly progressing disorder with short survival. The peak incidence of primary myelofibrosis is between 50 and 70 years of age. Citing the Committee Papers for 'ruxolitinib for disease-related splenomegaly or symptoms in adults with myelofibrosis' (review of TA289), the CS (Page 13) reports that Epidemiological estimates for myelofibrosis in UK patients suggest a prevalence of 2.2/100,000 and an incidence of 0.4/100,000.[1]

Diagnosis

There are two scoring systems available for diagnosing and stratifying myelofibrosis - the International Prognostic Scoring System (IPSS) and the Dynamic International Prognostic Scoring System (DIPSS) or DIPSS Plus. These are used to classify patients into one of four risk groups (low, intermediate-1, intermediate-2, and high-risk) based on factors such as age, presence of constitutional symptoms, and haematological parameters.[8]

2.2 Critique of company's overview of current service provision

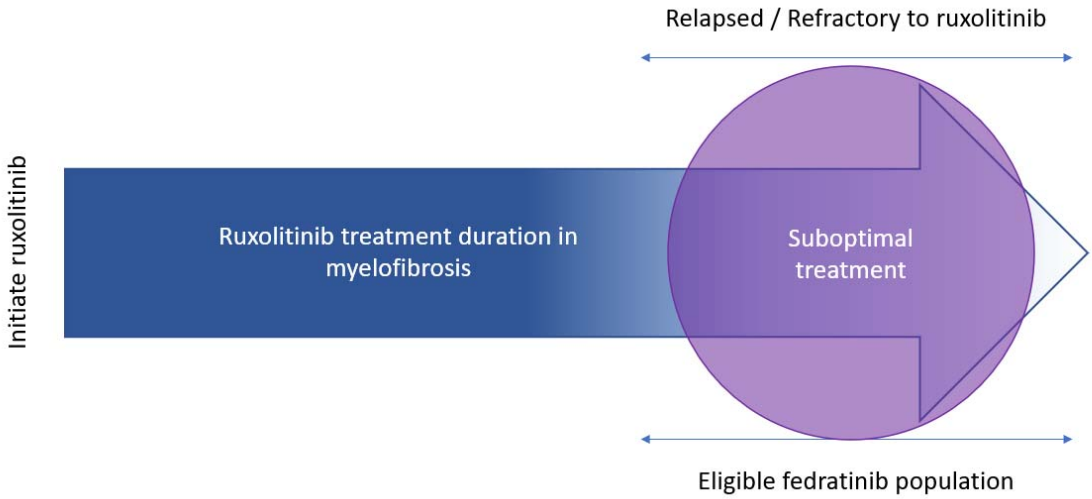
The ERG considers the company's overview of current service provision to be reasonable, in that the company acknowledges that: (i) allogeneic stem cell transplant (ASCT) is the only potentially curative treatment for myelofibrosis but, that ASCT is only suitable for people who are fit enough to undergo treatment; (ii) ruxolitinib is the only targeted treatment currently approved for myelofibrosis by the European Medicines Agency (EMA) and is used to improve disease-related splenomegaly or symptoms and prolong survival in patients ineligible for curative treatment with ASCT; (iii) currently, only ruxolitinib is recommended by NICE for use in patients classified as having intermediate-2 or high-risk disease (CS,[1] Page 17).

The company presents a clinical pathway of care for intermediate-2 and high-risk myelofibrosis patients in England. The ERG provides a brief summary of this in this section.

Proposed patient/treatment pathway

For people with MF, an advisory board to the company proposed that many patients continue to receive suboptimal treatment with ruxolitinib, despite being relapsed or refractory and that reasons for this include the lack of treatment options, concerns regarding the potential for a pro-inflammatory state, and deterioration due to ruxolitinib withdrawal (symptoms include acute relapse of disease symptoms, accelerated splenomegaly, worsening of cytopenias, and occasional haemodynamic decompensation (including a septic shock-like syndrome).[1] The proposed position in the treatment pathway reproduced from the CS,[1] is presented in Figure 1.

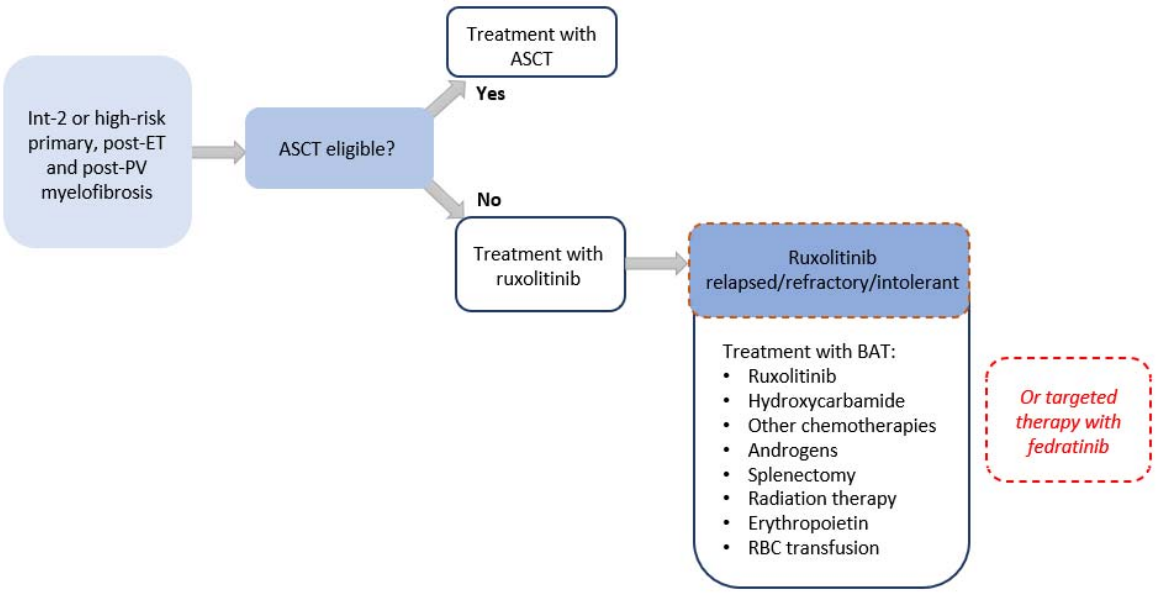
Figure 1: Schematic representation of the proposed position in the treatment pathway presented in the CS (Figure 3)



Reproduced from the CS, Page 18.[1]

The CS reports that best available therapy (BAT), including suboptimal ruxolitinib, includes treatment options that are largely supportive and do not significantly alter the course of the disease (treatments such as hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin and red blood cell transfusion) (CS,[1] Page 18). The CS also comments that suboptimal ruxolitinib in UK clinical practice aligns with observations from the PERSIST-2 (45%[9]) and SIMPLIFY-2 (89%[10]) RCTs used by the company in the ITC, where considerable proportions of patients in the BAT arms were receiving ruxolitinib. The CS also reports that data on survival in patients who continue suboptimal ruxolitinib is uncertain (CS,[1] Page 18). The clinical pathway of care for patients with myelofibrosis in England (source not reported in the CS), and potential position of fedratinib within this pathway, reproduced from the CS,[1] is presented in Figure 2.

Figure 2: Clinical pathway of care for intermediate-2 and high-risk myelofibrosis patients in England presented in the CS (Figure 4)



Reproduced from the CS Page 19.[1]

Clinical advice received by the ERG agreed with the pathway of care, but with some variation on what is used in BAT in the UK due to cost and/or adverse effects, e.g., androgens and erythropoietin not being used.

The CS comments that, given that there are currently no disease-modifying treatment options available to UK patients who are no longer responding to ruxolitinib, the introduction of fedratinib to the pathway of care would provide an opportunity for targeted therapy in a patient population otherwise associated with poor survival outcomes (Page 19).[1]

2.3 Critique of company's definition of the decision problem

2.3.1 Population

The target population in the company's decision problem is 'Adults with primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis that have been treated with ruxolitinib'.[1] This is different to the broader population in the NICE scope which is 'Adults with primary myelofibrosis, post polycythaemia vera myelofibrosis or post essential thrombocythaemia myelofibrosis'.[2] The CS focuses on fedratinib (INREBIC®, Celgene) as a second-line treatment in patients who have either relapsed on ruxolitinib, are refractory on ruxolitinib, or intolerant of ruxolitinib (JAKAVI®)'.[1]

The key clinical evidence submitted by the company is derived from a single-arm, open-label, non-randomised, phase 2, multicentre study (JAKARTA-2); of fedratinib in myelofibrosis patients who were currently or previously treated with ruxolitinib for at least 28 days and who had symptomatic intermediate-1 risk, intermediate-2 or high-risk disease.[11] Patients had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib and therefore were not currently treated with ruxolitinib nor were they previously treated for at least 28 days. Intermediate-2 or high-risk myelofibrosis is associated with a poor overall prognosis and limited survival time.[12] JAKARTA-2 reflects a slightly broader demographic than the target population in the CS, given that it includes 16 patients with intermediate-1 disease.[1] ruxolitinib is recommended by NICE for patients with intermediate-2 or high-risk disease.[13]

The JAKARTA-2 study recruited adult patients with a current diagnosis of intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis, or post-essential thrombocythaemia myelofibrosis, found to be ruxolitinib resistant or intolerant after at least 14 days of treatment and was undertaken in Austria, Canada, France, Germany, Italy, the Netherlands, Spain, the UK, and the USA. Clinical advice received by the ERG suggested that the population in this study is generally comparable to the UK myelofibrosis population. The CS also presented evidence from the JAKARTA study, a Phase III, double-blind, placebo-controlled RCT of 289 patients with intermediate-2 or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis.[14] This RCT did not contribute to the ITC or the company's model and, as such reference to this study is only made where relevant in this ERG report.

2.3.2 Intervention

The intervention evaluated in the CS is fedratinib (INREBIC®, Celgene), an oral kinase inhibitor with activity against wild-type and mutationally activated JAK2.[1] fedratinib is a JAK2-selective inhibitor with higher inhibitory activity for JAK2 over family members JAK1, JAK3 and TYK2. Fedratinib inhibits cytokine induced STAT3 phosphorylation in whole blood from myelofibrosis patients. A single dose administration of 300, 400, or 500 mg of fedratinib resulted in maximal inhibition of STAT3 phosphorylation approximately 2 hours after dosing, with values returning to near baseline at 24 hours.[15] The intervention matches that in the NICE scope.[2]

Fedratinib (INREBIC®, Celgene) currently does not have a marketing authorisation in the UK for treating myelofibrosis.[2] The anticipated indication for fedratinib is: for the treatment of disease-related splenomegaly or symptoms in adult patients with primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis who are JAK inhibitor naïve or who have been treated with ruxolitinib.' fedratinib has not been studied in patients with platelets <50 x 109 /L at baseline and may not be appropriate for use in this population.[1] A marketing authorisation

application for this indication was submitted to the European Medicines Agency (EMA) in [redacted].[15] Ruxolitinib is the only targeted treatment currently approved for myelofibrosis by the European Medicines Agency (EMA)

The Summary of Product Characteristics (SmPC) reports that Fedratinib is contraindicated in patients who have hypersensitivity to the active substance or to any of its excipients.[15] Special warnings and precautions for use listed in the SmPC are: encephalopathy, including Wernicke's encephalopathy; anaemia and thrombocytopenia, gastrointestinal events, hepatic toxicity, elevated amylase/lipase, and product excipients.[15] The SmPC considers special populations to be patients with: renal impairment, hepatic impairment, the elderly, and paediatric populations.[15] The SmPC advice to females of reproductive potential is to: avoid becoming pregnant whilst receiving fedratinib and to use effective contraception during treatment, that there are no data from the use of fedratinib in pregnant women, that fedratinib is not recommended during pregnancy, that it is unknown whether fedratinib/metabolites are excreted in human milk, that women should not breastfeed during treatment with fedratinib; and that there are no human data on the effect of fedratinib on fertility.[15]

Adverse events (AEs) including laboratory abnormalities which occurred most frequently in > 20% of patients listed in the SmPC are: diarrhoea (56.4%), nausea (59.7%), anaemia (44.2%), vomiting (49.5%), fatigue (25.0%), thrombocytopenia (21.2%), and constipation (20.6%).[15]

Fedratinib is administered orally as a single daily dose of 400 mg (four 100 mg tablets) taken with or without food. The list price for fedratinib is £6,119.68 per pack (120 x 100 mg capsules). The Patient Access Scheme (PAS) simple discount net price is £[redacted]. The cost-effectiveness results presented by the company are based on the PAS price.

2.3.3 Comparators

For patients with no previous treatment with ruxolitinib, the comparators in the NICE scope are ruxolitinib (for people with intermediate-2 risk or high-risk disease) and established clinical practice (including but not limited to hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin and red blood cell [RBC] transfusion). For patients with no previous treatment with ruxolitinib or in who ruxolitinib is not appropriate, the comparator in the NICE scope is established clinical practice only.[2]

The comparator in the CS is established clinical practice, otherwise referred to as BAT (including but not limited to ruxolitinib, hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin, and RBC transfusion).[1] The CS reports that the use of BAT as a comparator aligns with the design of comparative clinical trials in MF and previous ruxolitinib economic modelling

(Page 98).[1] Clinical advice received by the ERG confirmed that BAT described in the CS is comparable to UK clinical practice.

As the comparator in the CS is BAT and the JAKARTA-2 study did not include a comparison to BAT,[11] the company included two randomised controlled trials (RCTs)[9, 10] comparing other Janus kinase-2 (JAK2) inhibitors to BAT to undertake an unanchored indirect treatment comparison (ITC) of fedratinib compared to BAT. PERSIST-2 was a Phase III, international, open-label RCT in patients with myelofibrosis of pacritinib compared to BAT,[9] and SIMPLIFY-2 was a Phase III, international, open-label RCT in patients with myelofibrosis of momelotinib to with BAT.[10]

2.3.4 Outcomes

The outcomes in the decision problem in the CS are:

- Spleen size
- Symptom relief (including itch, pain and fatigue)
- Overall survival
- Response rate
- Haematological parameters (including RBC transfusion and blood count)
- Adverse effects of treatment
- Health-related quality of life (HRQoL).[1]

These outcomes match those in the NICE scope, with the exception of not including progression-free survival (PFS).[2] The CS did not include PFS as "*there is no standardised definition of progression in myelofibrosis and, therefore, it is not a measure used in any clinical trials*" (CS,[1] Page 9). Clinical advice received by the ERG on this was equivocal, with some advice suggesting that splenomegaly, becoming cytopenic, evidence of more blasts in peripheral blood, or bone marrow biopsy results; are indicative of disease progression.

Spleen response rate, was assessed in the JAKARTA-2 study as the proportion of patients with a ≥ 35% spleen volume reduction (SVR) at end of treatment cycle six (EOC6) relative to baseline, as measured by MRI/CT scan. Secondary outcomes included: the proportion of patients with a ≥ 35% SVR at EOC3; duration of spleen response as measured by MRI/CT; spleen volume and percent change of spleen volume at EOC3 and EOC6 from baseline as measured by MRI/CT; proportion of patients with a ≥ 50% reduction in spleen size by palpation at EOC3 and EOC6, relative to baseline; symptom response rate, defined as the proportion of patients with ≥ 50% reduction in the TSS at EOC6 relative to baseline. The key exploratory assessments were overall survival (OS), defined as the proportion of patients alive at the time of final analysis; and change in HRQoL using EORTC QLQ-C30 V3.0.[11]

Outcomes of ≥ 35% SVR from baseline to Week 24/EOC6 and ≥ 50% reduction in TSS from baseline to 24 weeks were also assessed in the two RCTs included in the ITC.[9, 10]

2.3.5 Other relevant factors

Equity

No equity issue was raised in the CS.[1]

Adherence and treatment continuation

Adherence to treatment is not reported in the CS.[1] The CS describes the mechanism of action for fedratinib as selectively inhibiting JAK2, with higher inhibitory activity for JAK2 over family members JAK1, JAK3 and TYK2; and that fedratinib is a more selective inhibitor of JAK2 than ruxolitinib which inhibits both subtypes, JAK1 and JAK2 (CS, Table 2). The ERG's clinical advisors stated that MF patients who relapse on ruxolitinib often continue with ruxolitinib at a suboptimal dose if it continues to provide symptom control, e.g., symptoms of sweating and fatigue.

Ongoing studies

The company searched appropriate sources to identify ongoing studies. The CS reports on two ongoing studies of fedratinib (CS, Page 72).[1] The Phase III, single-arm FREEDOM study of fedratinib in patients with intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib is currently recruiting and due to read out in 2022 (US study);[16] and the Phase III FREEDOM-2 study of fedratinib compared with BAT in patients with intermediate or high-risk primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis and previously treated with ruxolitinib is currently recruiting and due to read out in 2022.[17]

Patient Access Scheme

The CS reports a Patient Access Scheme for fedratinib as: [redacted] per 120 pack of 100mg tablets (CS, Table 2).[1]

3 CLINICAL EFFECTIVENESS

This section presents a review of the clinical effectiveness evidence reported in the CS[1] for fedratinib for splenomegaly and symptoms in myelofibrosis (CS Section B.2.).

3.1 Critique of the methods of review(s)

The company performed a systematic literature review (SLR) to identify primary intervention trials (randomised controlled trials [RCTs] and prospective non-RCTs) assessing the efficacy and safety of fedratinib or comparator therapies in patients with myelofibrosis. The CS (Appendix D, Page 14) notes that the study inclusion eligibility criteria are broader than the population in the decision problem of the CS, and included studies in patients both previously exposed to, and naïve to, JAK inhibitors. In addition, a broader set of interventions were included than relevant to the UK setting.

The clinical evidence provided in the CS comprises a single-arm study (JAKARTA-2) of fedratinib in myelofibrosis patients who were currently or previously treated with ruxolitinib, (CS, Section B.2.1 to Section B.2.7), and two RCTs of JAK2 inhibitors to compared to BAT (PERSIST-2, pacratinib;[9] SIMPLIFY-2, momelotinib;[10]) that were used to undertake an ITC of fedratinib compared to BAT (CS, Section 2.9).[1] Patients in JAKARTA-2 had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib and therefore were not currently treated with ruxolitinib.

3.1.1 Searches

For the original searches, several electronic bibliographic databases were searched in August 2018 including: MEDLINE, MEDLINE In-Process, Embase, Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), and Database of Abstracts of Reviews. The company performed one clinical effectiveness search to identify all clinical effectiveness and safety studies of fedratinib or comparator treatments of patients who have myelofibrosis, and searched several electronic bibliographic databases in February 2020 (Appendix D.1 Identification and selection of relevant studies): MEDLINE [via Embase.com], MEDLINE in Process [via PubMed.com], EMBASE [via Embase.com], Cochrane Database of Systematic Reviews [via Wiley], Cochrane Central Register of Controlled Trials [via Wiley], and Database of Abstracts of Reviews of Effects [via Wiley].

The ERG is unclear regarding the origin of the applied study design filter used in the CS as, to date, there has been no published all-in-one search filter that is able to retrieve RCTs, non-randomised controlled trials and real-world evidence (including retrospective and prospective observational studies). As such, the ERG cannot confirm the robustness of company's applied study filter in capturing relevant evidence.

The company also applied an English language limit to the MEDLINE and Embase searches. It is recommended that all eligible studies should be identified and assessed regardless of language even if it is considered that the exclusion of non-English publications does not change the conclusions of systematic reviews.

In Appendix D Tables 2-4, pages 5-14 of the CS, the company restricted the keyword searching of intervention and comparator terms to title and abstract fields only and excluded drug trade names in the strategies. The company's response to clarification question 3 on literature searching regarding this states that the intervention and comparator terms were searched using a mix of both free-text search terms and Emtree/MeSH terms which removes the possibility of missing relevant studies. Whilst MeSH and free-text terms have been searched, there is no published evidence to show that restricting the search to title and abstract field searching or the omission of drug trade names is sufficiently comprehensive to retrieve all eligible studies. The ERG recommends multi-purpose field searching in Embase which include searches in over ten fields including drug trade name and drug manufacturer.

The company also searched the clinicaltrials.gov trials registry although the company did not report on the search terms used and list of included studies from the trial's registry search. Supplementary searches by the company include searching several key conference abstract websites in the last three years. The company also searched several key HTA websites for previous technology submissions.

The CS (Section B.3.3.3) states that the August 2018 iteration of the clinical SLR was updated systematically using Embase to identify overall survival evidence for patients after discontinuation of ruxolitinib. However, it was unclear from the CS whether this search was separate and differed from that in the CS Appendix D.1.1.3. The company's response to clarification question 1 (priority question) on literature searching regarding this[18] states that the "*targeted review*" (the ERG considers this unclear, but may refer to the search for OS evidence) included searches of Embase until February 2019, and together these sources retrieved 4,011 publications.

In Appendix D of the CS, the company reports the full literature search strategies for clinical effectiveness; however, the cost-effectiveness evidence searches were not presented in the CS Appendices G (cost-effectiveness studies), H (HRQoL studies), or I (Cost and healthcare resources). The company's response to clarification question 2 on literature searching regarding this[18] included a copy of these searches.

3.1.2 In Appendix D Tables 2-4, pages 5-14 of the CS, the company also restricted the keyword searching of intervention and comparator terms to title and abstract fields only and excluded drug trade names in the strategies. The company's response to clarification question 3 on literature searching regarding this[18] states that the intervention and comparator terms were searched using a mix of both free-text search terms and Emtree/MeSH terms which removes the possibility of missing relevant studies. The ERG notes that searching both MeSH and free-text terms is standard practice in systematic review. However, the ERG considers that, as the CS evaluates drugs that have both trade and generic names, that there is no published evidence to show that the CS search is comprehensive, given that the company limited the search to title and abstract field searching or omitting the trade names. The ERG can therefore not guarantee that key studies have not been missed by the company. Inclusion criteria

The inclusion and exclusion criteria for the systematic review reported in the CS[1] are in accordance with the NICE scope,[2] with the exception that the population in the CS decision problem was patients who have either relapsed on ruxolitinib, are refractory to ruxolitinib, or intolerant of ruxolitinib. The company notes that "*the eligibility criteria were broader than the population addressed in the decision problem of this submission and included studies in patients both previously exposed to, and naïve to, Janus kinase (JAK) inhibitors. In addition, a broader set of interventions were included than relevant to the UK setting*" (CS, Appendix D, Page 14).

A copy of the inclusion and exclusion criteria, reproduced from the CS[2] are presented in

Table 3.

Table 3: Inclusion and exclusion criteria in systematic review search strategy (reproduced from Table 11 of the CS, Appendix D)

Category	Inclusion criteria	Exclusion criteria
Population	• Adult patients • Patients with intermediate-1, intermediate-2 and high-risk myelofibrosis (including primary, PPV-MF, or PET-MF), or myelofibrosis of indeterminate/undescribed risk	• Patients with low-risk myelofibrosis • Healthy volunteers • Children only (<18 years)
Interventions	• Anagrelide • Azacytidine • Cytarabine • Danazol • Darbepoetin alpha • Decitabine • Epoetin alpha • Epoetin beta • Fedratinib • Flucytosine • Guadecitabine • Hydroxycarbamide • Interferon • Lenalidomide • Melphalan • Mercaptopurine • Momelotinib • Pacritinib • Prednisolone • Prednisone • Pomalidomide • Ruxolitinib • Thalidomide • Thioguanine • Zebularine • Non-pharmacological interventions (such as allo-SCT)	• Studies assessing interventions not on the list
Comparators	• Placebo • Best supportive care • Any other pharmacological agents • Splenectomy • Non-pharmacological interventions (such as allo-SCT)	• No restrictions

Category	Inclusion criteria	Exclusion criteria
Outcomes	The data extraction was done in the Excel®-based extraction template shared and agreed with Celgene. Some of the outcomes were: • Spleen volume • Total symptom score (from any instrument) • Overall survival • Progression-free survival • Leukaemia-free survival • Patient-reported outcomes • Safety • Tolerability Subgroups • Age • Region • Baseline platelet counts • Patients with/without prior JAKi exposure • Primary/secondary myelofibrosis • Prognostic score (intermediate-1, intermediate-2, high-risk/intermediate-2, high-risk)	• Pharmacokinetics • Economic outcomes
Study type	• RCTs • Clinical trials (non-RCTs and single arm) • Prospective observational studies • Retrospective studies • Cross-sectional studies • Systematic reviews[a]	• Letters, comments, and editorials • Non-systematic reviews • Case reports and case series • Pre-clinical trials and animal experiments • Publications with redundant information
Time limit	• Original SLR: Data inception to 20 August 2018 • SLR Update 1: 1 August 2018 to 4 October 2019 • SLR Update 2: 1 September 2019 to 29 February 2020	No limit
Language	English only	Non-English

Key: allo-SCT, allogenic stem cell transplantation; JAKi, Janus kinase inhibitor; MF, myelofibrosis; non-RCT, non-randomized controlled trial; PET, post-essential thrombocythemia; PPV, post-polycythaemia vera; RCT, randomized controlled trial.

Appendix D.1.2 of the CS reports that the citation sifting stage was undertaken by a single reviewer, but did not report if any secondary independent checking (considered systematic review best practice) of either all records or a proportion was undertaken. Appendix D.1.2 of the CS reports that studies (the ERG assumes at the full-text stage, although this is not clear) were assessed for eligibility by two independent reviewers, with disagreements adjudicated by a third reviewer. However, it is not clear if the reviewers worked collaboratively or independently (systematic review best practice). The

company's response to clarification question C1 regarding this[18] states that "*studies were assessed for eligibility by two independent reviewers, with disagreements adjudicated by a third reviewer. This was applied at both title/abstract and full-text screening stage to ensure everything is quality checked*" The ERG considers this to be an adequate screening and study selection method for a systematic review.

3.1.3 Critique of data extraction

Details regarding the company's data extraction methods are reported in the CS, Appendix D.1.3.2.[1]

The CS (Appendix D), states that data were then extracted into a data extraction template by one reviewer and validated by a second reviewer. The company's response to clarification question C3 regarding if data extraction was undertaken independently and how many items were double-checked[18] states that "*All the extracted data was quality checked independently by a second reviewer*" The ERG considers this to be an adequate data extraction method for a systematic review.

Data extracted by the company in the CS from JAKARTA-2 were adults aged ≥18 years who had a current diagnosis of primary myelofibrosis, post-polycythaemia vera myelofibrosis, or post-essential study[11] are reported in Section 3.2.2 of this ERG report, and data extracted by the company from the PERSIST-2[9] and SIMPLIFY-2[10] RCTs (used by the company in the ITC), are reported in Section 4.3.5 of this ERG report. All data were checked against the published trial reports[9-11] and the JAKARTA-2 CSR[19] by the ERG.

3.1.4 Quality assessment

Quality assessment of the JAKARTA-2 study[11] is presented in Appendix D.3., Table 23, of the CS.[1] Quality assessment of this study was undertaken by the company using the Downs and Black checklist.[20] The ERG considers this an appropriate quality assessment method for single-arm intervention studies such as the JAKARTA-2 study.[11] A critique of the quality assessment of the PERSIST-2[9] and SIMPLIFY-2[10] RCTs, used by the company in the ITC, is presented in Section 4.3.4 of this ERG report.

3.1.5 Evidence synthesis

The company presented a narrative synthesis of the evidence for fedratinib for splenomegaly and symptoms in myelofibrosis. The ERG considers the narrative synthesis approach undertaken by the company to be acceptable. In addition, the company provided the following justification for not undertaking a network meta-analysis (CS, Page 53): "*As a single study (JAKARTA-2) provides data for fedratinib in patients treated with ruxolitinib, meta-analysis of intervention studies is not required.*

"*An ITC has been conducted to demonstrate the comparative efficacy and safety of fedratinib versus BAT, and is described in detail in Section B.2.9."*

The company's systematic review identified three studies that investigated either fedratinib or BAT in a patient population that had received prior JAK-inhibitor treatment; JAKARTA-2, PERSIST-2 and SIMPLIFY-2. These trials were included by the company in the ITC as they investigated SVR and/or TSS reduction and could therefore be compared with evidence for fedratinib from JAKARTA-2 (CS, Page 53).

Further details of the ERG's critique of the ITC are presented in Section 3.4 of this ERG report.

3.2 Critique of trials of the technology of interest, their analysis and interpretation

3.2.1 Included trials of fedratinib for splenomegaly and symptoms in myelofibrosis

The company identified one single-arm study of fedratinib for splenomegaly and symptoms in myelofibrosis. which was considered relevant to the decision problem (JAKARTA-2[11]). In JAKARTA-2, patients received oral fedratinib at a starting dose of 400 mg once per day, for six consecutive 28-day cycles. Dose adjustments of 100 mg/day were allowed to a minimum of 200 mg/day (due to toxicity) and a maximum of 600 mg/day (if the patient had not achieved a 50% reduction in spleen size by palpation and unacceptable toxicity had not been reported.[11]

*s*Eligibility criteria of the JAKARTA-2 study*

Inclusion criteria for JAKARTA-2[11] were adults aged ≥18 years who had a current diagnosis of primary myelofibrosis, post-polycythaemia vera myelofibrosis, or post-essential thrombocythaemia myelofibrosis according to the 2008 WHO classifications,[21] of intermediate-1, intermediate-2, or high-risk disease (according to the Dynamic International Prognostic Scoring System).[22] Risk categorisation was carried out using the IPSS or DIPSS in patients enrolled after Protocol Amendment 3 (CS Page 23). Patients with intermediate-1 disease had to have constitutional symptoms. Patients were also required to have received ruxolitinib therapy for the treatment of primary myelofibrosis, post-polycythaemia vera myelofibrosis, or post-essential thrombocythaemia myelofibrosis for at least 14 days (unless the patient discontinued due to intolerance or allergy within 14 days). Patients also had to have palpable splenomegaly (≥5 cm below the left costal margin), Eastern Cooperative Oncology Group performance status of 2 or less, and life expectancy of 6 months or more.

Patients who had received chemotherapy, including ruxolitinib, within 14 days before the start of the study (except hydroxyurea, which was permitted within 1 day of initiation of fedratinib), a history of other malignancies, and platelet count of less than 50×10^9 platelets per L, were excluded.[11]

Characteristics of the JAKARTA-2 study

Details of study location, treatments and numbers randomised, prohibited and disallowed medications and, primary and other outcomes reported in the CS are presented in Table 4 (reproduced from CS Table 4).

JAKARTA- 2 was a single-arm, open-label, non-randomised, phase 2, multicentre study, undertaken at 42 sites[1] across: Austria, Canada, France, Germany, Italy, the Netherlands, Spain, the UK, and the USA.[11] Treatment was oral fedratinib at a starting dose of 400 mg once per day, for six consecutive 28-day cycles. Dose adjustments of 100 mg/day were allowed to a minimum of 200 mg/day (due to toxicity) and a maximum of 600 mg/day (if the patient had not achieved a 50% reduction in spleen size by palpation and unacceptable toxicity had not been reported).[11]

The ERG has some concerns with the design of the JAKARTA-2 study that are covered in section 3.2.5 of this ERG report.

Patients were defined as resistant or intolerant to ruxolitinib by investigator assessment.[19]

Treatment with cytotoxic or immunosuppressive therapy, including hydroxycarbamide or systemic corticosteroids (i.e. > 10 mg/day prednisone or equivalent for > 5 days) was prohibited as was any other investigational agents during the study (CS,[1] Table 4). Other medications were not to be used prior to inclusion (CS,[1] Table 4) (see Table 4).

In JAKARTA-2 there was a screening period of up to 28 days, followed by a treatment phase of six 28-day cycles of fedratinib (24 weeks) and a follow-up visit (approximately 30 days following the last dose of fedratinib). There was no stopping rule and patients could remain on fedratinib until disease progression (enlargement of spleen volume) of ≥25% compared with baseline) or unacceptable toxicity.[11]

Table 4: Study location, treatments and numbers randomised, prohibited and disallowed medications and, primary and other outcomes for fedratinib from the JAKARTA-2 study relevant to the decision problem (reproduced from Table 4 of the CS)

Trial number (acronym)	NCT01523171 (JAKARTA-2)
Location	JAKARTA-2 was conducted in 42 sites in nine countries, including one site in the UK
Eligibility criteria for patients	Key inclusion criteria: • Patients who previously received ruxolitinib therapy for the treatment of primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis for at least 14 days (unless the patient discontinued due to intolerance or allergy within 14 days) • Palpable splenomegaly (≥ 5 cm below the left costal margin) • ECOG Performance Status of 2 or less, and life expectancy of 6 months or more Key exclusion criteria: • Received any chemotherapy, including ruxolitinib, within 14 days before the start of the study (except hydroxycarbamide, which was permitted within 1 day of initiation of fedratinib) • A history of other malignancies • Platelet count of $< 50 \times 10^9$ /L
Settings and locations where the data were collected	Steps taken to ensure the accuracy and reliability of the clinical study data included regular site monitoring visits to review study progress, investigator and patient compliance with the protocol requirements, and any emergent problems Data entry and validation were carried out using standard validated remote data capture computer software (Oracle Clinical RDC Version 4.6). Data were stored in an Oracle database on a UNIX server. Data entry was performed directly from the investigator site from the data source documents and signed electronically by the authorised site personnel. Any modification in the database was traced using an audit trail
Trial drugs	400 mg fedratinib was given orally, once daily. If there was a lack of adequate spleen response, the fedratinib dose could be titrated upwards in 100 mg/day increments up to a maximum of 600 mg/day Study treatment continued until disease progression or unacceptable toxicity
Permitted and disallowed concomitant medication	Patients could not receive any other drug treatment for their disease while on study. Treatment with cytotoxic or immunosuppressive therapy, including hydroxycarbamide or systemic corticosteroids (i.e. > 10 mg/day prednisone or equivalent for > 5 days) was prohibited. Use of any other investigational agents during the study was prohibited. The following medications were not to be used prior to inclusion: any chemotherapy, immunomodulatory drug therapy (e.g. thalidomide, interferon-α), anagrelide, immunosuppressive therapy, corticosteroids > 10 mg/day prednisone or equivalent, or growth factor treatment (e.g. erythropoietin), or hormones (e.g. androgens, danazol) within 14 days prior to initiation of fedratinib; and darbepoetin within 28 days prior to initiation of fedratinib

Primary outcome (including scoring methods and timings of assessments)	The primary outcome, spleen response rate, was defined as the proportion of patients with a ≥ 35% SVR at EOC6 relative to baseline, as measured by MRI/CT scan. The MRI/CT scans were reviewed by an independent central imaging laboratory, where reviewers were blinded to the fedratinib doses.
Other outcomes used in the economic model/specified in the scope	Secondary efficacy assessments: • Spleen response rate, defined as the proportion of patients with a ≥ 35% SVR at EOC3, relative to baseline, as measured by MRI/CT scan • Duration of spleen response as measured by MRI/CT • Spleen volume and percent change of spleen volume at EOC3 and EOC6 from baseline as measured by MRI/CT • Proportion of patients with a ≥ 50% reduction in spleen size by palpation at EOC3 and EOC6, relative to baseline • Symptom response rate, defined as the proportion of patients with ≥ 50% reduction in the TSS at EOC6 relative to baseline Key exploratory assessments: • OS, defined as the proportion of patients alive at the time of final analysis • Change in HRQoL using EORTC QLQ-C30 V3.0
Pre-planned subgroups	Analyses of spleen volume reduction and symptom response rate were measured in pre-planned subgroups of: • Demographic factors and baseline disease characteristics • Platelet count at baseline (< 100 x 10^9/L or ≥ 100 x 10^9/L) • Patients resistant versus intolerant to ruxolitinib
Key: CSR, clinical study report; CT, computed tomography; ECOG, Eastern Cooperative Oncology Group; EOC3, end of Cycle 3; EOC6, end of Cycle 6; HRQoL, health-related quality of life; MRI, magnetic resonance imaging; OS, overall survival; SVR, spleen volume reduction; TSS, total symptom score. Source: Harrison et al. 2017[11] and JAKARTA-2 CSR.[19]	

Sample size, power calculation, and intention-to-treat of the JAKARTA-2 study

The CS[1] (Page 34) reports, "*Assuming 25% of patients achieved the primary endpoint of ≥ 35% reduction in spleen volume from baseline, 70 evaluable patients were required to provide at least 90% power (at a one-sided 2.5% α level) to test the null hypothesis of ≤ 10% of patients achieving the primary endpoint.*". Company response to clarification question A2 regarding the justification for the sample size calculation and the response rate to rule out for patients treated with standard of care in the study population[18] states that "*assuming the primary endpoint (spleen volume reduction) was [redacted]%, then [redacted] evaluable patients were to provide at least [redacted]% power at a 1-sided [redacted]% α-level to test the null hypothesis of ≤[redacted]% response rate. Based on the COMFORT-I study results, approximately 60% of patients who received ruxolitinib were non-responders. Therefore, 60% of [redacted] evaluable patients (i.e., [redacted]) were to provide [redacted]% power to test a spleen response rate ≤[redacted]% for the subgroup of patients who did not reach the primary endpoint of spleen response during the ruxolitinib studies.*"

In JAKARTA-2, the intention-to-treat (ITT) population comprised all 97 patients enrolled in the study.[1] The primary analysis of JAKARTA-2 was conducted in the per protocol (PP) population (n = 83), patients with evaluable baseline and at least one post baseline MRI/CT scan of spleen volume [EOC3 or EOC6] and no important protocol deviations that could impact the efficacy outcome).[1] Missing data for patients who did not reach EOC6 owing to the clinical hold, were handled using the last observation carried forward (LOCF) method.[1]

The CS (Page 35) reports that, "*As the PP population represents a smaller population than the ITT population, and considering the statistical limitations of the LOCF method, the analyses in the PP population are considered supportive to the ITT population*."[1] The ERG notes that there are methods other than LOCF for dealing with missing data and that a per protocol analysis does not mitigate any limitations associated with LOCF. Furthermore, if both sets of results are biased but similar then this does not mean that either results are correct.

Baseline characteristics of study patients in JAKARTA-2

Details of patient baseline characteristics of the ITT population in the JAKARTA-2 study[11] presented in the CS[1] (Table 5) are presented in Table 5.

The proportion of patients who were female was 45% (44/97), and the median age in years (range) was 67 (38 to 83).

Myelofibrosis Disease types were: Primary myelofibrosis 53/97 (55%), Post-polycythaemia vera myelofibrosis, 25/97 (26%); Post-essential thrombocythaemia myelofibrosis, 19/97 (20%). Disease risk

status was: Intermediate-1, 16/97 (17%); Intermediate-2, 47/97 (48%); High-risk, 34/97 (35%). The ERG's clinical advisors believed that this disease type and risk status was generally comparable to the UK myelofibrosis population.

The CS (Page 29) reports that the JAKARTA-2 study reflects a slightly broader demographic than the target population in the NICE Scope,[2] given that it includes 16 (16.5%) patients with intermediate-1 disease, but that the company's analyses were adjusted to consider removal of these intermediate-1 patients.[1] The ERG's clinical advisors believed that this was reasonable, given that intermediate-1 patients would have better treatment outcomes.

The CS (Page 31) reports that patients in the JAKARTA-2 study were heavily pre-treated, with [redacted]% having received at least two prior anticancer therapies and [redacted]% having received at least four prior anticancer therapies. The CS also reports that all 97 patients enrolled in the JAKARTA-2 study had received prior treatment with ruxolitinib, with a median exposure of 10.7 months.[1] Besides ruxolitinib, the most common anticancer therapy was hydroxycarbamide, received by [redacted]% of patients in JAKARTA-2.[1] The ERG's clinical advisors believed that this was comparable to treatment in the UK setting.

Table 5: Baseline characteristics of patients in the JAKARTA-2 study relevant to the decision problem (reproduced from Table 5 of the (CS)

	Patients (N=97)
Median age, years (range)	67 (38, 83)
Sex, n (%)	
Male	53 (55%)
Female	44 (45%)
Race, n (%)[a]	
White	92 (94.8%)
Black	1 (1.0%)
Asian	4 (4.1%)
Median weight, kg (range)	73.0 (47.0, 105.7)
Disease type, n (%)	
Primary myelofibrosis	53 (55%)
Post-polycythaemia vera	25 (26%)
Post-essential thrombocythaemia	19 (20%)
Risk status, n (%)[b]	
Intermediate-1	16 (17%)
Intermediate-2	47 (48%)
High-risk	34 (35%)
Median time since diagnosis, years (range)	4.1 (0.3, 24.5)
JAK2 mutational profile, n (%)	
Wild-type	29 (30%)

	Patients (N=97)
Mutant	61 (63%)
Missing	7 (7%)
RBC transfusion dependence status, n (%)[c]	
Yes	14 (14%)
No	83 (86%)
Platelet count, n (%)	
<50 x 10^9/L	1 (1%)
≥50 x 10^9/L to <100 x 10^9/L	32 (33%)
≥100 x 10^9/L	64 (66%)
Haemoglobin level, n (%)	
<10 g/dL	51 (53%)
≥10 g/dL	46 (47%)
ECOG, n (%)	
0	26 (26.8%)
1	45 (46.4%)
2	23 (23.7%)
Missing	3 (3.1%)
Constitutional symptoms[d]	
Yes	93 (95.9%)
No	4 (4.1%)
Median baseline spleen volume, ml (range)	2894 (737, 7815)
Median baseline spleen size, cm (range)[e]	18 (5, 36)

Key: CT, computed tomography; DIPSS, Dynamic International Prognostic Scoring System; ECOG, Eastern Cooperative Oncology Group; IPSS, International Prognostic Scoring System; ITT, intention-to-treat; JAK2, Janus kinase 2; MPN-SAF, myeloproliferative neoplasm symptom assessment form; MRI, magnetic resonance imaging; RBC, red blood cell.
Notes: Spleen volume was measured by MRI/CT scan and reviewed in a blinded fashion by a central imaging laboratory. Spleen size was measured by palpation (i.e. length in cm). [a], the race categories in the electronic case report form were Caucasian/White, Black, Asian/Oriental and other. The race categories in this table were standardised for consistency across fedratinib clinical study reports. Race 'other' is not presented because there were no patients in the category; [b], risk category per IPSS or DIPSS for patients enrolled after Protocol Amendment 3; [c], receiving ≥ 2 units/month of RBC transfusions over 3 months prior to first dose; [d], a subject had constitutional symptoms if any of the symptoms in the baseline MPN-SAF (night sweats, itching, abdominal discomfort, abdominal pain, early satiety, bone pain) had a value greater than zero; [e], below lower coastal region.
Source: Harrison et al. 2017,[11] Harrison et al. 2019,[23] Harrison et al. 2020,[24] and JAKARTA-2 CSR.[19]

Patient flow in the JAKARTA-2 study, the 2013 clinical hold on fedratinib, and the change in criteria for ruxolitinib failure and JAKARTA-2 reanalysis

In the JAKARTA-2 study, one hundred twenty-seven patients were recruited of whom 97 were enrolled and received at least one dose of febratinib.[11]

In November 2013, JAKARTA-2 and other studies in the fedratinib clinical development program were placed on clinical hold by the US FDA following reports of suspected Wernicke's encephalopathy (WE), and patients receiving fedratinib at that time were discontinued from ongoing fedratinib treatment.[11] Sixty-three patients (65%) of the 97 patients in JAKARTA-2 discontinued treatment due to study termination following the fedratinib clinical hold.[24]

The CS (Page 25), reports that "*Based on experts' review, there was a consensus of a clear diagnosis of WE in one out of the eight suspected patients, with the other diagnoses remaining uncertain or inconclusive*." Furthermore, that WE can be can be mitigated with routine thiamine monitoring and thiamine replacement, and that the FDA lifted the clinical hold on fedratinib in 2017.
See next subheading for other reasons for discontinuation.

In the JAKARTA-2 study, patients were originally classified as resistant or intolerant to ruxolitinib per the investigators' assessments.[1] A reanalysis of the efficacy of fedratinib in JAKARTA-2 was subsequently performed on patients determined to be relapsed or refractory or intolerant to ruxolitinib, based on more stringent criteria:

- Relapsed: ruxolitinib treatment for ≥3 months with spleen regrowth, defined as <10% SVR or <30% decrease in spleen size from baseline, following an initial response.
- Refractory: ruxolitinib treatment for ≥3 months with <10% SVR or <30% decrease in spleen size from baseline.
- Intolerant: ruxolitinib treatment for ≥28 days complicated by development of RBC transfusion requirement (≥2 units per month for 2 months); or grade ≥3 thrombocytopenia, anaemia, hematoma and/or haemorrhage while receiving ruxolitinib.[24]

Based on these criteria, 79 patients were considered to be relapsed, refractory or intolerant on prior ruxolitinib therapy and comprised the Stringent Criteria Cohort.[24] The remaining 18 patients were excluded from the Stringent Criteria Cohort because they had an adequate response to ruxolitinib (n = 3), were missing ruxolitinib response data (n = 8), or did not receive ≥3 months of ruxolitinib treatment.[24]

A Sensitivity Cohort then comprised 66 patients who received six fedratinib treatment cycles or discontinued before EOC6 for reasons other than study being terminated by the sponsor.[24]

Patients completing the JAKARTA-2 study / included in the company's analysis
In the JAKARTA-2 study, all patients (N=97) discontinued treatment with fedratinib. Sixty-three (65%) discontinued due to the early termination of the study; 18 (19%) due to AEs, six (6%) due to disease progression, three (3%) because of patient decision and seven (7%) for other reasons.[1]

The CS (Page 35) reports that "*The ITT population comprised all 97 patients enrolled in the study and provides the largest sample size and statistically robust source for evaluations of efficacy in JAKARTA-2. A reanalysis of JAKARTA-2 data was conducted to confirm the efficacy of fedratinib in subsets of enrolled patients who met new stringent definitions of ruxolitinib relapsed, refractory or intolerant* "[1]

In the JAKARTA-2 study analysis, the last observation carried forward (LOCF) method was used to account for patients that did not meet EOC6 due to the clinical hold. In the updated analyses presented in the CS (full ITT population and reanalysis populations), LOCF was not applied. A patient without a Cycle 6 assessment was considered a non-responder.

Outcomes of the JAKARTA-2 study

The primary outcome in JAKARTA-2 was spleen response rate, defined as the proportion of patients with a ≥ 35% SVR at EOC6 relative to baseline, as measured by MRI/CT scan (CS,[1] Table 4). Secondary outcomes from JAKARTA-2, were:

- Spleen response rate, defined as the proportion of patients with a ≥ 35% SVR at EOC3, relative to baseline, as measured by MRI/CT scan
- Duration of spleen response as measured by MRI/CT
- Spleen volume and percent change of spleen volume at EOC3 and EOC6 from baseline as measured by MRI/CT
- Proportion of patients with a ≥ 50% reduction in spleen size by palpation at EOC3 and EOC6, relative to baseline
- Symptom response rate, defined as the proportion of patients with ≥ 50% reduction in the TSS at EOC6 relative to baseline (CS,[1] Table 4)

Key exploratory assessments used by the company in the economic model were:

- OS, defined as the proportion of patients alive at the time of final analysis
- Change in HRQoL using EORTC QLQ-C30 V3.0 (CS,[1] Table 4)

3.2.2 Efficacy results for trials of fedratinib for splenomegaly and symptoms in myelofibrosis

The company's overview of JAKARTA- 2 results

The CS (Page 38) presents an overview of the JAKARTA- 2 clinical effectiveness and reports that, in the ITT population, 31% of patients achieved the primary outcome of spleen response rate defined as ≥35% SVR at EOC6. In the ITT population, 25% of patients achieved the key secondary outcome of symptom response rate defined as ≥50% reduction in TSS at EOC6. An overview of SVR at EOC6, TSS at EOC6, and 12-month OS rate for the ITT population, intermediate-2 and high-risk patients, the reanalysis Stringent Criteria Cohort and the Reanalysis Sensitivity Cohort, is presented in

Table 6 (reproduced from Table 9 of the CS).

Table 6: Overview of fedratinib efficacy from JAKARTA-2 (adapted from CS Table 9)

Endpoint	Measure	JAKARTA-2 (Phase II, previously treated with ruxolitinib)			
		ITT population	Int-2 and high-risk patients[a]	Reanalysis: Stringent Criteria Cohort[b]	Reanalysis: Sensitivity Cohort[c]
		fedratinib 400 mg (n=97)	fedratinib 400 mg (n=81)	fedratinib 400 mg (n=79)	fedratinib 400 mg (n=66)
Spleen response rate	≥35% SVR at EOC6, % (95% CI)	31% (22, 41)	[redacted]	30% (21, 42)	36% (25, 49)
Symptom response rate[d]	≥50% reduction in TSS at EOC6, % (95% CI)	25% (17, 35)	[redacted]	27% (17, 39)	32% (21, 45)
Overall survival	12-month OS rate, % (95% CI)	[redacted]	[redacted]	NA	[redacted]
	HR versus placebo (95% CI)	NE	NE	NE	NE

Key: CI, confidence interval; CSR, clinical study report; EOC6, end of Cycle 6; HR, hazard ratio; int, intermediate; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; NA, not assessed; NE, not estimable; OS, overall survival; TSS, total symptom score.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold; [d], this outcome was assessed in the MF-SAF population which was defined as patients with evaluable baseline and ≥1 post-baseline MF-SAF assessment. For JAKARTA-2, this includes 90 of the ITT patients, 74 of the Stringent Criteria Cohort patients, and 62 of the Sensitivity Cohort patients. For JAKARTA, this includes 91 patients in the fedratinib groups and 85 patients in the placebo group; [e], symptom response rate in the Int-2/high-risk subgroup did not apply evaluable baseline and ≥1 post-baseline MF-SAF assessment criteria.
Source: Harrison et al. 2020,[24] Pardanani et al. 2015,[14] JAKARTA-2 CSR[19] and data on file.[25, 26]

Further details of these and other efficacy results from JAKARTA-2 are presented in this section below.

JAKARTA- 2 spleen response rate (≥ 35% SVR) at EOC6

Details of the primary outcome in the CS of spleen response rate (≥ 35% SVR) at EOC6 are presented in Table 7.

In the ITT population (N=97), 31% (n=30; 95%CI, 22% to 41%) achieved ≥ 35% SVR at EOC6. In the Int-2/high-risk subpopulation (N=81), [redacted]% ([redacted]; 95%CI, [redacted]%, [redacted]%) achieved ≥ 35% SVR at EOC6.

From the reanalysis (applying more stringent criteria of ruxolitinib relapse and intolerance to the ITT population), the results were comparable to the ITT population; with 30% (24/79) of Stringent Criteria Cohort patients demonstrating ≥ 35% SVR at EOC6 (95%CI, 21% to 42%).

From the analysis removing patients who were directly impacted by the clinical hold (i.e., the Sensitivity Cohort), 36% (24/66) of patients demonstrated SVR at EOC6 (95% CI, 25%, 49%).

Table 7: Details of Spleen response rates at EOC6 (≥ 35% SVR from the JAKARTA-2 study (adapted from the CS Table 10)

≥ 35% SVR at EOC6, n/N (%; 95% CI)	fedratinib 400 mg			
	ITT population (n = 97)	**Int-2/high-risk patients[a] (n = 81)**	**Reanalysis: Stringent Criteria Cohort[b] (n = 79)**	**Reanalysis: Sensitivity Cohort[c] (n = 66)**
	30/97 (31%; 95%CI, 22% to 41%)	[redacted]	24/79 (30%; 95% CI, 21% to 42%)	24/66 (36%; 95%CI, 25% to 49%)

Key: CI, confidence interval; EOC6, end of Cycle 6; int, intermediate; ITT, intent-to-treat; SVR, spleen volume reduction.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold.
Source: Harrison et al. 2020,[24] and data on file.[26]

JAKARTA- 2 spleen response rate (≥ 35% SVR) at EOC3

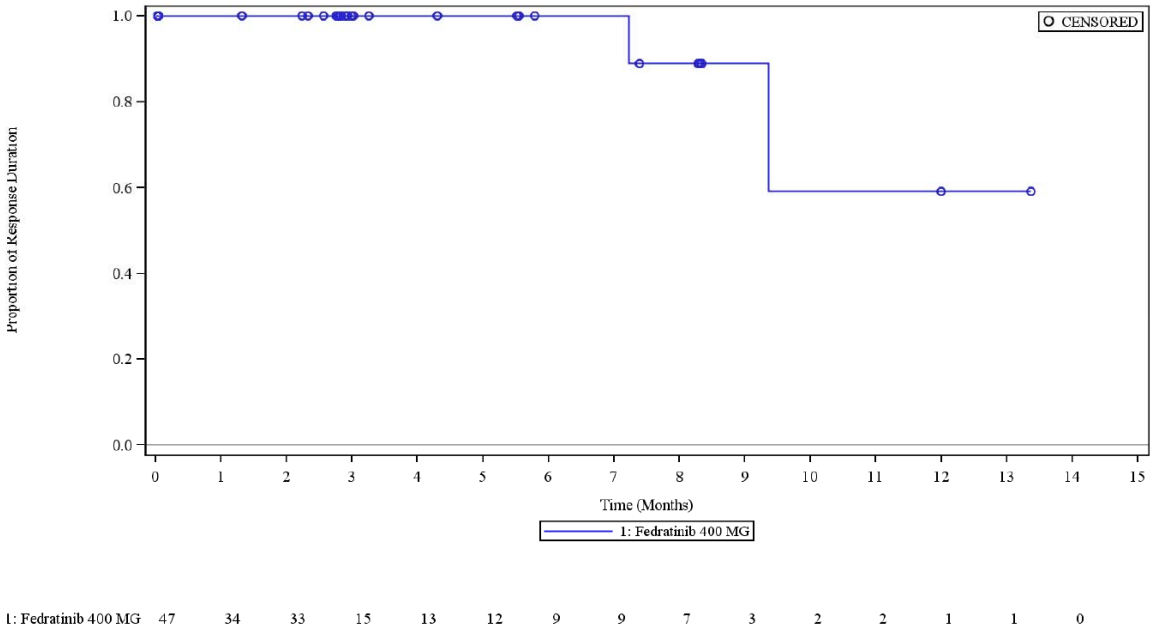
Details of the secondary outcome in the CS of spleen response rate (≥ 35% SVR) at EOC6 are reported in the CS (Page 41) as: 40% (95% CI: 30%, 51%) in the ITT population, 43% (95% CI: 32%, 55%) in the Stringent Criteria Cohort, and 41% (95% CI: 29%, 54%) in the Sensitivity Cohort.

JAKARTA- 2 duration of spleen response rate

For the secondary outcome of duration of response, The CS reports that in the analysis responders were all patients who at any time achieved ≥ 35% SVR from baseline that included 47 patients in JAKARTA-2. The CS reports that, based on Kaplan–Meier (KM) estimates, only 25% of patients had a duration of response of less than 9.4 months and the median duration was not reached (NR). Median spleen volume response duration was also NR (95% CI: 7.2 months, NR) in both the Stringent Criteria Cohort (n = 41 responders) and the Sensitivity Cohort (n = 34 responders). The CS (Page 41) notes that this outcome measure required extensive censoring due to early termination. A copy of the Kaplan–Meier plot of duration of spleen response reproduced from the CS, Figure 7 is presented in

Figure 3

Figure 3: Kaplan–Meier plot of duration of spleen response, ≥ 35 SVR at any time on study treatment (JAKARTA-2, ITT population, reproduced from CS Figure 7)

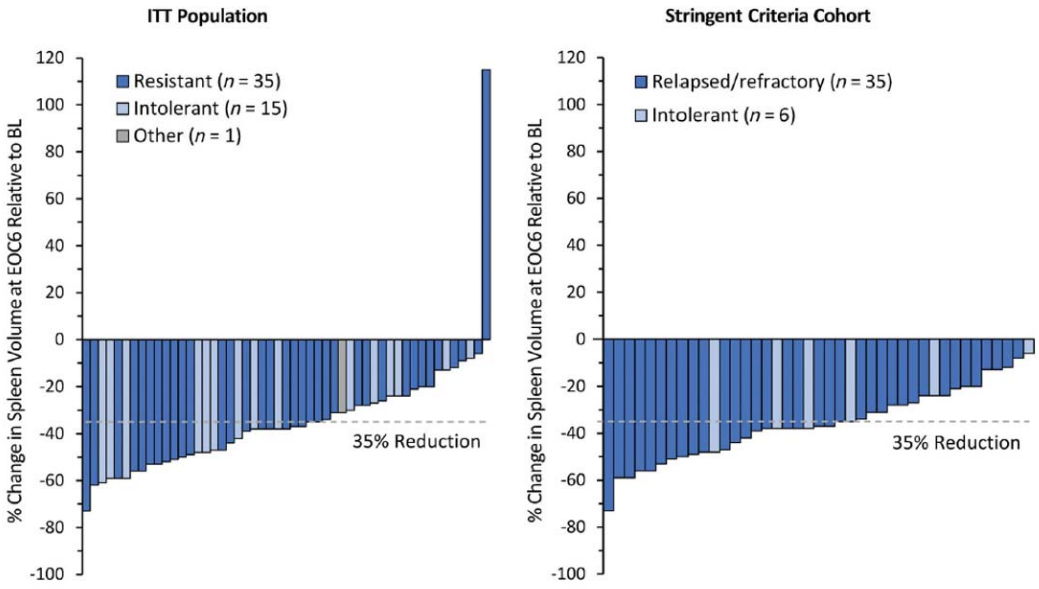


Key: ITT, intent-to-treat; SVR, spleen volume reduction.
Notes: patients at risk are shown along the horizontal axis. The duration of spleen response was calculated from the first date of spleen response (i.e. ≥ 35% SVR from baseline) to the first date of disease progression (i.e. ≥ 25% spleen volume increase from baseline) or death, whichever was earlier.
Source: Harrison et al. 2020.[24]

JAKARTA- 2 Percent change of spleen volume at EOC3 and EOC6

For the secondary outcome of percent change of spleen volume at EOC3 and EOC6, The CS reports that in the ITT population, the median percentage changes in spleen volume were [redacted]% at EOC3 (range: [redacted], [redacted]) and -38.0% at EOC6 (range: -73, -115), and that at EOC6, all patients except one in the ITT population and all patients in the Stringent Criteria Cohort showed a SVR (see Figure 4, reproduced from the CS Figure 8).

Figure 4: Individual changes in spleen volume from baseline to EOC6 (JAKARTA-2, ITT and Stringent Criteria Cohort, reproduced from CS Figure 8)



Key: BL, baseline; EOC6, end of Cycle 6; ITT, intent-to-treat.
Source: Harrison et al. 2020.[24]

JAKARTA-2 Spleen response rate by palpation at EOC3 and EOC6

For the secondary outcome of spleen response rate by palpation, the CS (Page 43) notes that this was defined as the proportion of patients with ≥ 50% reduction in spleen size.

The CS reports that in the ITT population, the proportion of patients with ≥50% reduction in spleen size were [redacted]% at EOC3 and 31% at EOC6 (Table 8, reproduced from CS Table 11).

The CS (Page 43) notes that the patients who demonstrated ≥35% SVR at EOC6 were the same patients who demonstrated ≥50% reduction in spleen size at EOC6.

Table 8: Spleen response rate by palpation (≥ 50% reduction in spleen size) at EOC3 and EOC6 (JAKARTA-2, ITT population, reproduced from Table 11 of the CS)

	fedratinib 400 mg (N=97)
EOC3	
n (%)	[redacted]
95% CI	[redacted]
EOC6	
n (%)	30 (31%)
95% CI	[redacted]

Key: CI, confidence interval; EOC3, end of Cycle 3; EOC6, end of Cycle 6; ITT, intent-to-treat.
Notes: Spleen size was measured by palpitation (i.e. length in cm)

Source: JAKARTA-2 CSR,[19] and Harrison 2020.[24]

The CS (Page 43) reports that results in the Stringent Criteria Cohort and Sensitivity Cohort were consistent with the ITT population, with reduction in spleen size of ≥ 50% at EOC6 observed in 30 (31%) and 24 (36%) patients, respectively.

JAKARTA-2 Symptom response rate (≥ 50% reduction in TSS) at EOC6

For the secondary outcome of Symptom response rate (≥ 50% reduction in TSS) at EOC6, the CS (Page 43) reports that the analyses of symptom response rate were performed using the MF-SAF Analysis Population, defined as patients with an evaluable baseline assessment of modified MF-SAF TSS, and at least one post-baseline evaluable assessment, and that symptom response rates were defined as the proportion of patients with ≥50% reduction in TSS from baseline to EOC6.

The CS (Page 43) reports that most evaluable patients demonstrated an improvement in TSS and more than a quarter of patients achieved a clinically meaningful threshold for response of ≥ 50% reduction. The proportion of patients in the MF-SAF Analysis Population with a ≥ 50% reduction in TSS at EOC6 was 27% (95% CI: 18%, 37%). Among patients with evaluable TSS data at baseline and EOC6, 82% reported some decrease in symptom severity with fedratinib.

Symptom response rates in the Stringent Criteria and Sensitivity Cohorts EOC6 were 27% (95% CI: 17 to 39) and 32% (95% CI: 21 to 45), respectively (Table 9, adapted from Table 12 of the CS).

The CS (Page 44) reports that in order to derive the most conservative plausible estimate, and to ensure comparability with reporting in other trials, the economic modelling for fedratinib calls upon results for symptom response rate in the ITT population. In this population, the proportion of patients with ≥ 50% reduction in TSS at EOC6 was [redacted]% ([redacted]/97). The results for the Int-2/high-risk subgroup of patients were consistent with the ITT population with [redacted]% (95% CI: [redacted]) achieving ≥ 50% reduction in TSS at EOC6 ([redacted]/81).

Table 9: Symptom response rates at EOC6 (≥ 50% TSS; JAKARTA-2, reproduced from CS Table 12)

≥ 50% reduction in TSS at EOC6	fedratinib 400 mg			
	All enrolled	**Int-2/high-risk patients[a]**	**Reanalysis: Stringent Criteria Cohort[b]**	**Reanalysis: Sensitivity Cohort[c]**
MF-SAF, N[d]	90	NA	74	62
% (95% CI)	27% (18, 37)	NA	27% (17, 39)	32% (21, 45)
ITT, N	97	81	NA	NA
n, % (95% CI)	[redacted]	[redacted]	NA	NA

Key: CI, confidence interval; CSR, clinical study report; EOC6, end of Cycle 6; int, intermediate; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; NA, not assessed.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold; [d], includes patients with evaluable baseline and ≥ 1 post-baseline MF-SAF assessment.
Source: Harrison et al. 2020,[24] and data on file.[26]

JAKARTA-2 Total symptom score by key symptoms

The CS (Page 45) reports that all key symptoms assessed in the MF-SAF Analysis Population in JAKARTA-2 showed an improvement at EOC6 in half of the evaluable patients, with median percent changes of:

- -83% in pain under ribs on left side
- -76% in night sweats
- -51% in early satiety
- -46% in abdominal discomfort
- -44% in pruritus
- -22% in bone or muscle pain

JAKARTA-2 Key exploratory outcome measures - Spleen or symptom response rate at EOC6

The CS (Page 45) reports that a combined endpoint of spleen or symptom response was strongly recommended as a modelling input by experts at an advisory board to the company, with the rationale that this outcome would be reflective of UK clinical practice given that the two track together.[27] Spleen or symptom response rate is defined in the CS (Page 45) as the number of patients achieving either ≥ 35% SVR or ≥ 50% reduction in TSS.

The CS (Page 45) reports that in the ITT population (N=97), [redacted]% ([redacted]; 95%CI, [redacted]% to [redacted]%) achieved ≥ 35% SVR or ≥ 50% reduction in TSS at EOC6. In the Int-2/high-risk subpopulation (N=81), [redacted]% (n=40; 95%CI, [redacted]%, [redacted]%) achieved this outcome. From the reanalysis (applying more stringent criteria of

ruxolitinib relapse and intolerance to the ITT population), the results were [redacted]% ([redacted]/79; 95%CI, [redacted]%, [redacted]%)) for the Stringent Criteria Cohort patients. From the analysis removing patients who were directly impacted by the clinical hold (i.e., the Sensitivity Cohort), [redacted]% ([redacted]/66; 95%CI, [redacted]%, [redacted]%) achieved this outcome. These results are presented in Table 10 (reproduced from Table 13 of the CS).

The CS (Page 45) reports that in JAKARTA-2, treatment with fedratinib was associated with almost half of patients achieving a spleen or symptom response rate at EOC6, with generally consistent results in ITT, Int-2/high-risk, and reanalysis cohorts (Table 10).

Table 10: Spleen or symptom response rates at EOC6 (≥ 35% SVR or ≥ 50% reduction in TSS; JAKARTA-2, reproduced from Table 13 of the CS)

≥ 35% SVR or ≥ 50% reduction in TSS at EOC6, n, % (95% CI)	fedratinib 400 mg			
	ITT population (n=97)	Int-2/high-risk patients[a] (n=81)	Reanalysis: Stringent Criteria Cohort[b] (n=79)	Reanalysis: Sensitivity Cohort[c] (n=66)
	[redacted]	[redacted]	[redacted]	[redacted]

Key: CI, confidence interval; CSR, clinical study report; EOC6, end of Cycle 6; int, intermediate; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; NA, not assessed.
Notes: [a], ITT population of JAKARTA-2 minus the 16 Int-1 patients; [b], reanalysis of ITT data in the ruxolitinib failure cohort defined using new stringent definitions of ruxolitinib relapsed/refractory; [c], the sensitivity cohort estimates fedratinib response without the impact of the clinical hold; [d], includes patients with evaluable baseline and ≥1 post-baseline MF-SAF assessment.
Source: Data on file.[26]

JAKARTA-2 Key exploratory outcome measures – Overall survival

The CS (Page 195) report that the OS data for JAKARTA-2 are immature and heavily censored owing to early study termination. At the time of the final analysis there were a total of [redacted] deaths; [redacted] deaths occurred whilst on-treatment and [redacted] deaths occurred more than 30-days post treatment. The proportion of patients alive at 12 months was [redacted]% (

Figure 5 reproduced from CS Figure 9).

Figure 5: Kaplan–Meier estimates of OS (JAKARTA-2, ITT population, reproduced from CS Figure 9)

Key: ITT, intent-to-treat; NC, not calculable; OS, overall survival.
Notes: OS is defined as the time interval from the date of first dose to the date of death due to any cause. In the absence of the confirmation of death, OS is censored at the last date patient was known to be alive.
Source: Data on file.[25]

The CS (Page 47) also reported that OS estimates for the Int-2/high-risk disease population were consistent with the ITT population, with [redacted]% patients alive at 12 months (Figure 4 reproduced from CS Figure 10).

Figure 6: Kaplan–Meier estimates of OS (JAKARTA-2, Int-2/high-risk population, reproduced from CS Figure 10)

Key: Int, intermediate; OS, overall survival.
Notes: OS is defined as the time interval from the date of first dose to the date of death due to any cause. In the absence of the confirmation of death, OS is censored at the last date patient was known to be alive.
Source: Data on file.[26]

JAKARTA-2 Quality of life – EORTC QLQ-C30

The CS (Page 48) defined the EORTC QLQ-C30 analysis population as all treated patients who had a baseline and ≥ 1 post-baseline assessment of the QLQ-C30 questionnaire(n = 90), and reported that completion rates for patients in the ITT population for each cycle ranged from % to % for all cycles.

The CS (Page 49) reports that at EOC6, mean changes from baseline in QLQ-C30 functional domain scores were:

- GHS QoL –
- Physical functioning domain – Role functioning domain – Social functioning domain –

For symptom domain scores, mean change in QLQ-C30 score from baseline to EOC6 were observed for appetite loss ( ), insomnia ( ), dyspnoea ( ), financial difficulties ( ), fatigue ( ) and pain ( ).

The mean changes from baseline in EORTC QLQ-C30 functional and symptom scores are presented in Figure 7 (reproduced from CS Figure 11) and Figure 8 (reproduced from CS Figure 12), respectively.

Figure 7: Mean change from baseline in QLQ-C30 functional scores (JAKARTA-2, EORTC QLQ-C30 Analysis Population, reproduced from CS Figure 11)

Key: EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Core Quality of Life.
Source: JAKARTA-2 CSR.[19]

Figure 8: Mean change from baseline in QLQ-C30 symptom scores (JAKARTA-2, EORTC QLQ-C30 Analysis Population, reproduced from CS Figure 12)

Key: EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Core Quality of Life.
Source: JAKARTA-2 CSR.[19]

JAKARTA-2 MF-SAF, myelofibrosis symptom assessment form

The CS (Page 44) reports that the proportion of patients in the MF-SAF Analysis Population with a ≥ 50% reduction in TSS at EOC6 was 27% (95% CI: 18%, 37%). Among patients with evaluable TSS data at baseline and EOC6, 82% reported some decrease in symptom severity with fedratinib.

The CS (Page 44) reports that symptom response rates in the Stringent Criteria and Sensitivity Cohorts were 27% (95% CI: 17, 39) and 32% (95% CI: 21, 45), respectively.

The CS (Page 44) reports that in the ITT population the proportion of patients with ≥ 50% reduction in TSS at EOC6 was [redacted]% ([redacted]). The results for the Int-2/high-risk subgroup of patients were consistent with the ITT population with [redacted] achieving ≥ 50% reduction in TSS at EOC6 (23/81).

JAKARTA-2 Subgroup analysis

The CS (Page 51) reports on subgroup analyses that were carried out to determine the treatment effect of fedratinib on clinically important subpopulations. The subgroups were: patients with platelet count at baseline of ≥ 50 x 10^9/L to < 100 x 10^9/L; patients with platelet count at baseline of ≥ 100 x 10^9/L; patients who were resistant to ruxolitinib; and patients in the ITT population who were intolerant of ruxolitinib.

In these subgroups the outcomes that were analysed were: ≥ 35% SVR at EOC6, ≥ 50% reduction in TSS at EOC3, and ≥ 50% reduction in TSS at EOC6.

The spleen response outcomes by platelet count subgroups are presented in

Table 11 (reproduced from CS Table 14) and the spleen response outcomes by the ruxolitinib resistant and intolerant subgroups are presented in Table 12 (reproduced from CS Table 15).

Table 11: Efficacy of fedratinib 400 mg by platelet count at baseline (JAKARTA-2, reproduced from CS Table 14)

	Platelet count at baseline	
	≥ 50 x 10^9/L to < 100 x 10^9/L	≥ 100 x 10^9/L
≥ 35% SVR at EOC6[a]		
ITT population, n/N	[redacted]/33	[redacted]/64
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC3[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC6[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]

Key: CI, confidence interval; CT, computed tomography; EOC3, end of Cycle 3; EOC6, end of Cycle 6; MF-SAF, myelofibrosis symptom assessment form; MRI, magnetic resonance imaging; SVR, spleen volume reduction; TSS, total symptom response.
Notes: [a], spleen volume was measured by MRI/CT scan and reviewed in a blinded fashion by a central imaging laboratory; [b], CI estimated using Clopper–Pearson Exact method; [c], TSS was defined as the sum of the daily average score of the six-item measures in a week: night sweats, pruritus, abdominal discomfort, early satiety, pain under ribs on left side and bone or muscle pain. For this analysis, patients with a baseline TSS equal to 0 are excluded (due to no place for symptom reduction). Patients with a missing TSS at the EOC6 were considered as non-responders.
Source: Harrison 2020,[24] and JAKARTA-2 CSR.[19]

Table 12: Efficacy of fedratinib 400 mg in patients resistant or intolerant to ruxolitinib at baseline (JAKARTA-2, ITT population, reproduced from CS Table 15)

	Resistant	Intolerant
≥ 35% SVR at EOC6[a]		
ITT population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC3[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]
≥ 50% reduction in TSS at EOC6[c]		
MF-SAF population, n/N	[redacted]	[redacted]
% (95% CI)[b]	[redacted]	[redacted]

Key: CI, confidence interval; EOC3, end of Cycle 3; EOC6, end of Cycle 6; ITT, intent-to-treat; MF-SAF, myelofibrosis symptom assessment form; SVR, spleen volume reduction; TSS, total symptom score.
Notes: Investigators' assessments of patients resistant or intolerant to ruxolitinib. One patient was neither resistant nor intolerant per investigator's assessment and was categorised under 'other: lack of efficacy'. [a], Spleen volume was measured by MRI/CT scan and reviewed in a blinded fashion by a central imaging laboratory; [b], CI estimated using Clopper–Pearson Exact method; [c], TSS was defined as the sum of the daily average score of the six-item measures in a week: night sweats, pruritus, abdominal discomfort, early satiety, pain under ribs on left side and bone or muscle pain.

	Resistant	Intolerant
For this analysis, patients with a baseline TSS equal to 0 are excluded (due to no place for symptom reduction). Patients with a missing TSS at the EOC6 were considered as non-responders. **Source:** Harrison 2020[24], and JAKARTA-2 CSR.[28]		

3.2.3 Safety results for trials of fedratinib for splenomegaly and symptoms in myelofibrosis

JAKARTA-2 Treatment exposure

The CS (Page 64) reports that in the all treated population (N=97), the median number of treatment cycles was six (inter quartile range 3.9–8.9). Fourteen out of 97 (14.4%) patients received more than 12 cycles. Treatment was discontinued due to early study termination in 63 of 97 (65%) patients. The remainder of patients discontinued study treatment due to AEs (19%), disease progression (6%), patient decision (3%), or other reasons (7%). Thirty-eight of 97 (39%) patients had at least one dose reduction, 13 (13%) had two dose reductions and four (4%) had more than two dose reductions. A total of 25 of 97 (25.8%) patients had a dose interruption for at least 7 consecutive days.

Most patients ([redacted]%) received the maximum daily dose of 400 mg fedratinib and almost all patients received ≥ 80% of the intended dose ([redacted]%).

JAKARTA-2 Summary safety data

The CS (Page 65) reports that the safety analyses were performed in the all treated population; defined as enrolled patients who took at least one dose (even if partial) of study medication (n=97).

The CS (Page 65) reports that all 97 patients had at least one treatment-emergent adverse event (TEAE) of any grade.[24, 28] Grade 3 or 4 TEAEs were reported by [redacted] (63%) patients, including transfusion dependency in [redacted] ([redacted]%) patients. Treatment-emergent serious adverse events (SAEs) were reported by [redacted] (34%) patients. Seven (7%) patients had a TEAE that led to death during treatment or follow-up; in four cases, the cause of death was determined to be due to disease progression and the other three cases were due to a TEAE considered not related to study treatment. TEAEs leading to treatment discontinuation occurred in [redacted] (20%) patients and TEAEs leading to dose modification occurred in [redacted] ([redacted]%) patients.

An overview of the TEAEs associated with fedratinib in JAKARTA-2 is provided in

Table 13 (reproduced from CS Table 22).

Table 13: Safety overview (JAKARTA-2, all treated population, reproduced from CS Table 22)

n (%)	fedratinib 400 mg (N = 97)
TEAE	97 (100%)
Treatment-related TEAE	[redacted]
Grade 3 or 4 TEAEs	61 (63%)
Treatment-related Grade 3 or 4 TEAEs	[redacted]
TEAE leading to death	7 (7%)
Treatment-related TEAE leading to death	0
Treatment-emergent SAEs	33 (34%)
Treatment-related treatment-emergent SAEs	[redacted]
TEAEs leading to permanent treatment discontinuation	19 (20%)
TEAEs leading to dose modification	[redacted]

Key: SAE, serious adverse event; TEAE, treatment-emergent adverse event.
Notes: Data are for patients with ≥ 1 TEAE.
Source: Harrison 2020,[24] and JAKARTA-2 CSR.[19]

JAKARTA-2 Common adverse event data

The CS (Page 66) reports that most common non-haematological TEAEs were gastrointestinal disorders including diarrhoea in 60/97 (62%) patients, nausea in 54/97 (56%) patients, vomiting in 40/97 (41%) patients, constipation in 20/97 (21%) patients, and abdominal pain in 12/97 (12%) patients. Other common non-haematological TEAEs in other system order classes included pruritus in 17/97 (17.5%) patients, fatigue in 15/97 (15.5%) patients, cough and headache in 13/97 (13%) patients each, urinary tract infection and dyspnoea in 12/97 (12%) patients each and dizziness in 11/97 (11%) patients.

The CS (Page 67) reports that the most common haematological TEAEs were anaemia in 47/97 (48%) patients and thrombocytopenia in 26/97 patients (27%). Grade 3 or 4 anaemia was reported in 37/97 (38%) patients and thrombocytopenia in 21/97 (22%) patients.

JAKARTA-2 Treatment-emergent SAEs

The CS (Page 68) reports that treatment-emergent SAEs were reported in 33/97 (34%) patients.[11, 28] The most common SAE was cardiac disorders, reported in five of 97 patients (5%). Pneumonia was reported in four of 97 patients (4%), pleural effusion in three (3%) and fall in [redacted] patients ([redacted]).

[redacted] patients ([redacted]%) had SAEs considered treatment related. Pneumonia was the only treatment-related SAE reported in more than one patient and occurred in [redacted] patients.

JAKARTA-2 Adverse events leading to treatment discontinuation

The CS (Page 68) reports that TEAEs leading to treatment discontinuation occurred in [redacted] (20%) patients, of whom [redacted] ([redacted]%) had a Grade 3 or 4 event. The most common reason for treatment discontinuation was Grade 3 or 4 thrombocytopenia, which occurred in two patients. [redacted] patient had disease transformation to AML, which was considered an AE, but the reason for discontinuation was recorded as disease progression.

One case of Grade 3 encephalopathy was reported, it was subsequently determined by an independent expert safety panel to be related to hepatic encephalopathy and inconsistent with WE. The event resolved within one week after discontinuation of fedratinib treatment.

JAKARTA-2 Adverse events leading to death

The CS (Page 70) reports that seven of 97 (7%) patients died during treatment in JAKARTA-2, but none of the deaths were deemed to be related to fedratinib. Three patients died due to fatal TEAEs of pneumonia, shock and cardiorespiratory arrest. The four other patients died due to disease progression as the main cause of death.

JAKARTA-2 Safety overview

The CS (Page 72) reports that most common TEAEs observed in JAKARTA-2 were consistent with the known safety profile of fedratinib, could be managed with dose modifications and were not a frequent reason for discontinuation of fedratinib. The most frequent Grade 3 or 4 events in JAKARTA-2 were anaemia and thrombocytopenia. The three fatal TEAEs (pneumonia, cardio-respiratory arrest and shock) were not considered to be related to fedratinib treatment. Analysis of the signs and symptoms that may be associated with events of WE in JAKARTA-2 were not suggestive of any confirmed cases. Furthermore, that increased clinical awareness of the potential for developing WE coupled with routine thiamine monitoring and thiamine replacement, sufficiently minimises the risk of developing this AE

3.2.4 Quality assessment results for trials of fedratinib for splenomegaly and symptoms in myelofibrosis

The CS (Appendix D, Page 19) states that quality assessment was undertaken using the Downs and Black Checklist.[20] The ERG considers this quality checklist to be appropriate for the JAKARTA-2 study design.

The results of the company's Downs and Black quality assessment of the JAKARTA-2 study are presented in Table 14 (adapted from CS, Appendix D Table 23). The ERG considers the company's application of the Downs and Black quality checklist to the JAKARTA-2 study to be accurate.

Table 14: Quality assessment of JAKARTA-2, using the Downs and Black checklist (adapted from CS, Appendix D Table 23)

Quality assessment item	CS response
1. Is the hypothesis/aim/objective of the study clearly described?	Yes
2. Are the main outcomes to be measured clearly described in the introduction or methods section?	Yes
3. Are the characteristics of the patients included in the study clearly described?	Yes
4. Are the interventions of interest clearly described?	Yes
5. Are the distributions of principal confounders in each group of patients to be compared clearly described?	No
6. Are the main findings of the study clearly described?	Yes
7. Does the study provide estimates of the random variability in the data for the main outcomes?	Yes
8. Have all important adverse events that may be a consequence of the intervention been reported?	Yes
9. Have the characteristics of patients lost to follow-up been described?	Yes
10. Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	No
11. Were the patients asked to participate in the study representative of the entire population from which they were recruited?	Yes
12. Were those patients who were prepared to participate representative of the entire population from which they were recruited?	Yes
13. Were the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive?	Yes
14. Was an attempt made to blind study patients to the intervention they have received?	N/A
15. Was an attempt made to blind those measuring the main outcomes of the intervention?	N/A
16. If any of the results of the study were based on 'data dredging', was this made clear?	No
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	No
18. Were the statistical tests used to assess the main outcomes appropriate?	Yes
19. Was compliance with the intervention(s) reliable?	Yes
20. Were the main outcome measures used accurate (valid and reliable)?	Yes
21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	N/A
22. Were study patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	N/A
23. Were study patients randomised to intervention groups?	N/A
24. Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	N/A
25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	Yes

Quality assessment item	CS response
26. Were losses of patients to follow-up considered?	Yes
Notes: CS, Company Submission; N/A, not applicable. **Source:** SLR report[29, 30]	

3.2.5 ERG critique of trials of the technology of interest

The ERG has some concerns with the design of the JAKARTA-2 study and using evidence from it to make reimbursement decisions in the target population:

- Only those Phase 2 studies with good results lead to treatments being evaluated in Phase 3 studies. Consequently, estimates of treatment effects from Phase 2 studies generally overestimate true treatment effects.[31]
- Populations defined by Phase 2 inclusion/exclusion criteria are often less diverse than in Phase 3 studies and tend to generate larger treatment effects.[31]
- The lack of a concurrent control means that the study is likely to suffer from the phenomenon known as regression to the mean such that recruitment to the study is a consequence of extreme values that return to their average values post-treatment even if there is no treatment effect.
- It is not clear why the study did not include a concurrent control. The established clinical practice for patients treated with ruxolitinib in the UK includes a basket of treatment options that are supportive but do not alter the course of disease.
- Although the primary outcome measure may be considered clinically relevant, dichotomising a continuous variable is statistically inefficient.

Some of the limitations associated with the design of the study can be mitigated by adjusting for all relevant prognostic factors and treatment effect modifiers. However, as discussed in Section 3.4 of this ERG report, it is impossible to specify the correct model and there will be residual bias.

Care should be taken when interpreting the duration of response as the results only include patients who were SVR responders. The ERG was unable to verify the statement made in the CS (Page 41) that, "*Based on Kaplan–Meier (KM) estimates, only 25% of patients had a duration of response of less than 9.4 months …*"

Pre-specified subgroup analyses were conducted of patients with a platelet count of between $\geq 50 \times 10^9$/L and $< 100 \times 10^9$/L or $\geq 100 \times 10^9$/L at baseline, and patients resistant and intolerant to ruxolitinib. Age was categorised as ≤ 65 and > 65 years and as ≤ 75 and > 75 years. Categorisation of continuous variables leads to a loss of information and implies that the effect of a covariate changes abruptly at the cut-offs rather than with each unit increase in the continuous variable. Continuous

variables should be modelled as continuous variables. The CS (Appendix E Page 40) claims that, "*results of the subgroup analyses of spleen response rate and symptom response rate in JAKARTA-2 were consistent across baseline demographic and disease characteristics subgroups, supporting the robustness of the results of the primary analysis.*" While the ERG considers it is not true that response rates were consistent across subgroups, no evidence is available to assess whether the relative effect of fedratinib compared to BAT varies according to baseline characteristics.

Overall Survival

To support the assertion that that there may be an OS benefit for fedratinib in patients treated with ruxolitinib, the CS (Page 48) claims that, "*JAKARTA demonstrated an OS benefit for fedratinib 400 mg versus placebo, with [redacted]% of patients alive at 12 months (OS hazard ratio [HR] [redacted]; 95% CI: [redacted], [redacted]; p=[redacted]).*" While the observed effect was favourable to fedratinib, there was uncertainty regarding the true treatment effect.

3.3 Critique of trials identified and included in the indirect treatment comparison

Details of the identification and methodology of the PERSIST-2 RCT[9] and the SIMPLIFY-2 RCT[10]), used by the company in the ITC analysis are described below.

2.3.1 Search Strategy

The company (CS, Page 53) undertook a systematic literature review (SLR) was conducted to identify evidence of relevance to the efficacy and safety of treatments for myelofibrosis which is critiqued in Section 3.1.1 of this ERG report.

The CS (Page 53) states that the searches identified three studies that investigated either fedratinib or BAT in a patient population that had received prior JAK-inhibitor treatment; JAKARTA-2,[11] PERSIST-2[9] and SIMPLIFY-2,[10] and that these trials were included as they investigated SVR and/or TSS reduction and could therefore be compared with evidence for fedratinib from JAKARTA-2.[11]

2.3.2 Study selection criteria

The CS (Page 53) states that studies in the ITC were included as they investigated SVR and/or TSS reduction and could therefore be compared with evidence for fedratinib from JAKARTA-2 The CS (Page 54) states that the ITT populations from the included studies represents the most appropriate populations for comparative purposes. The ERG acknowledges that an ITT population is generally preferred in a randomised controlled trial. However, the issue when making unanchored indirect comparisons between treatments is whether it is possible to adjust for difference between studies in patient characteristics.

The PRISMA flow diagram presented in Appendix D of the CS is unclear regarding how many of the studies identified as potentially relevant for inclusion in the ITC, were excluded at the full-text stage. In response to a request for clarification from the ERG, the company provided a table of 326 studies that were excluded at the full-text stage according to the CS PRISMA flow diagram. However, the reasons for exclusion were not included and it was also unclear which of these had been identified as potentially relevant for inclusion in the ITC.

2.3.3 Studies identified

The CS (Consilient Health 2019)[32] (Page 53) states that three studies that investigated either fedratinib or BAT in a patient population that had received prior JAK-inhibitor treatment were included in the ITC: JAKARTA-2,[11] PERSIST-2[9] and SIMPLIFY-2.[10] The CS (Page 53) states that that these studies were included as they investigated SVR and/or TSS reduction and could therefore be compared with evidence for fedratinib from JAKARTA-2. Details of these studies are presented in Table 15 (reproduced from CS Table 16).

Table 15: Summary of the studies used in the indirect treatment comparison (reproduced from CS Table 16)

	JAKARTA-2	PERSIST-2	SIMPLIFY-2 [a]
Phase	II	III	III
Design	Single-arm	RCT	RCT
Method of blinding	Open-label	Open-label	Open-label
Intervention (N)	fedratinib 400 mg, once daily (starting dose) (97 [ITT])	Pacritinib 400 mg, once daily (75 [ITT]) and pacritinib 200 mg, twice daily (74 [ITT])	Momelotinib 200 mg once daily (104 [ITT])
Comparator	NA	BAT (72 [ITT efficacy population]): • Ruxolitinib (45%) • Watch and wait (19%) • Hydroxycarbamide (hydroxyurea)(19%) • Prednisone (13%) • Danazol (5%) • Thalidomide (3%) • Decitabine (2%) • Interferon-alpha (2%)	BAT (52 [ITT]): • Ruxolitinib (89%) • Hydroxycarbamide (hydroxyurea) (23%) • Corticosteroids (12%)
Location	Multicentre	Multicentre	Multicentre

	JAKARTA-2	PERSIST-2	SIMPLIFY-2 [a]
Method of randomisation	NA	1:1:1 ratio stratified by geographic region, risk category and rebound platelet count	2:1 stratified by transfusion dependence and by baseline TSS
Crossover	NA	After Week 24 or progression of splenomegaly before Week 24	After completion of the randomized phase (24 weeks), all subjects were eligible to receive momelotinib in an extended treatment phase
Key inclusion/exclusion criteria			
Prior JAK inhibitor treatment	Prior ruxolitinib (≥14 days of exposure or <14 days if patients discontinued ruxolitinib due to intolerability or allergy)	Prior treatment with one or two other JAK-inhibitors was allowed, patients could be JAK-inhibitor naïve: • 33 patients had prior ruxolitinib (45.8%) • 39 patients were ruxolitinib naïve (54.2%)	Currently or previously treated with ruxolitinib (≥28 days) and either: • RBC transfusion needed while on ruxolitinib • Dose adjustment of ruxolitinib to <20 mg twice daily and Grade 3 thrombocytopenia/ anaemia/hematoma
Platelet count	≥50 x 10^9/L	≤100 x 10^9/L	There were no inclusion/exclusion criteria for platelet count at baseline
Diagnosis	PMF, PPV-MF, PET-MF	PMF, PPV-MF, PET-MF	PMF, PPV-MF, PET-MF
DIPSS[b]	• Intermediate-1 with symptoms • Intermediate-2 • High-risk	• Intermediate-1 • Intermediate-2 • High-risk	• Intermediate-1 with symptomatic splenomegaly/ hepatomegaly • Intermediate-2 • High-risk

Key: BAT, best available therapy; DIPSS, Dynamic International Prognostic Scoring System; ITC, indirect treatment comparison; ITT, intent-to-treat; L, litre; N, number of subjects; NA, not applicable; PET-MF, post-essential thrombocytopenia myelofibrosis; PMF, primary myelofibrosis; PPV-MF, post-polycythaemia vera myelofibrosis; RBC, red blood cell; RCT, randomised controlled trial; TSS, total symptom score.
Notes: [a], only the most frequent treatments received were reported; the percentages in this table do not sum to 100% as patients could have received more than one therapy; [b], DIPSS score calculation: 1 point for each of the following criteria: age >65 years, white cell count ≥25 x 10^9/L, haemoglobin < 10 g/dL, peripheral blood blasts ≥1%, constitutional symptoms (weight loss and/or unexplained fever or excessive sweats).
Source: ITC report.[33]

2.3.4 Quality assessment of studies included in the ITCs

Details of the CS quality assessment of the JAKARTA-2 study are presented in Section 3.1.4 of this ERG report.

Quality assessment of the PERSIST-2[9] and SIMPLIFY-2[10] RCTs, used by the company in the ITC, is presented in Appendix D.1.4.3., Table 21, of the CS.[1] The CS reports that quality assessment of this study was undertaken by the company "*based on the NICE-recommended checklist for bias*" (CS,[1] Page 37). The company's response to clarification question C4 regarding the supporting citation for this[18] states that the NICE checklist from the NICE STA user guide was used.[34] The ERG considers this an appropriate quality assessment method for RCTs.

The results of the company's NICE recommended checklist for bias assessment of the PERSIST-2[9] and SIMPLIFY-2[10] RCTs, are presented in Table 14 (reproduced from CS, Appendix D Table 21). The ERG considers the company's quality assessment of these RCTs to be accurate.

Table 16: Quality assessment, SIMPLIFY 2 and PERSIST-2 (reproduced from CS Table 21)

Author and year of publication	SIMPLIFY 2	PERSIST-2
Was randomisation carried out appropriately?	Yes	Yes
Was the concealment of treatment allocation adequate?	Yes	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	No	No
Were the care providers, patients, and outcome assessors blind to treatment allocation?	No	No
Were there any unexpected imbalances in dropouts between groups?	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	Yes	Yes
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes
Source: SLR report[30]		

2.3.5 Critique of studies included in the ITC

For details of the JAKARTA-2 study, please see Section 3.2 of this ERG report. The ERG's critique of the PERSIST-2[9] and SIMPLIFY-2[10] RCTs are presented below.

Study designs of studies in the ITC

Both the PERSIST-2[9] and SIMPLIFY-2[10] RCTs were Phase III, international multicenter, open label trials. PERSIST-2 was conducted in Australia, Belgium, Canada, Czechia, France, Germany, Hungary, Netherlands, New Zealand, Russian Federation, the UK, and the USA (n centres not reported). SIMPLIFY-2 was conducted at 52 centres in Canada, France, Germany, the UK, and the USA.

Population characteristics of studies in the ITC

Eligibility criteria of PERSIST-2 and SIMPLIFY-2 are presented in Appendix D, Table 13 of the CS.

Eligibility criteria of PERSIST-2 were: JAK-inhibitor naïve or prior treatment with one or two other JAK-inhibitors, a platelet count of $\leq 100 \times 10^9$/L, a myelofibrosis diagnosis of PMF, PPV-MF, or PET-MF; a DIPSS score of Intermediate-1, Intermediate-2, or high-risk; an ECOG PS of 0, 1, 2, or 3; and palpable spleen of ≥5 cm.

Eligibility criteria of SIMPLIFY-2 were: patients who had currently or previously been treated with ruxolitinib (at least 28 days) and either RBC transfusion needed while on ruxolitinib, or a dose adjustment of ruxolitinib to <20 mg twice daily and Grade 3 thrombocytopenia/ anaemia/haematoma; a myelofibrosis diagnosis of PMF, PPV-MF, or PET-MF; a DIPSS score of Intermediate-1 with symptomatic splenomegaly/ hepatomegaly, Intermediate-2, or High-risk; an ECOG PS of 0, 1, or 2; and palpable spleen ≥ 5 cm. There were no eligibility criteria for platelet count at baseline for SIMPLIFY-2.

For comparative purposes, the eligibility criteria for JAKARTA-2 (from Table 4 of the CS), and SIMPLIFY-2 and PERSIST-2 (from Table 13 of the CS Appendix D) are presented in

Table 17.

Table 17: Patient eligibility criteria in JAKARTA-2, PERSIST-2 and SIMPLIFY-2 (from CS Table 4 and CS Appendix D Table 13)

	JAKARTA-2	PERSIST-2	SIMPLIFY-2
	Source: Harrison et al. 2017[11] and JAKARTA-2 CSR.[19]	**Source:** ITC report[33]	
Prior JAK-inhibitor treatment	Prior treatment with ruxolitinib therapy for the treatment of for at least 14 days (unless the patient discontinued due to intolerance or allergy within 14 days) defined as resistant or intolerant to ruxolitinib by investigator assessment.	Prior treatment with one or two other JAK-inhibitors was allowed (patients could be JAK-inhibitor naïve)	Currently or previously treated with RUX (at least 28 days) and either: RBC transfusion needed while on RUX, or Dose adjustment of RUX to < 20 mg twice daily and Grade 3 thrombocytopenia/ anaemia/haematoma
Platelet count	$<50 \times 10^9$/L	$\leq 100 \times 10^9$/L	There was no inclusion/exclusion criteria for platelet count at baseline
Diagnosis	primary myelofibrosis, post-polycythaemia vera myelofibrosis or post-essential thrombocythaemia myelofibrosis	PMF, PPV-MF, PET-MF	PMF, PPV-MF, PET-MF
DIPSS[a]	Intermediate-1 Intermediate-2 High-risk	Intermediate-1 Intermediate-2 High-risk	Intermediate-1 with symptomatic splenomegaly/ hepatomegaly Intermediate-2 High-risk
ECOG PS	0, 1, 2	0, 1, 2, 3	0, 1, 2
Palpable spleen ≥ 5 cm	Yes	Yes	Yes

Key: BAT, best available therapy; ECOG PS, Eastern Cooperative Oncology Group Performance Status; DIPSS, Dynamic International Prognostic Scoring System; JAK, Janus kinase; L, litre; MF, myelofibrosis; PET-MF, post-essential thrombocythemia-myelofibrosis; PMF, primary myelofibrosis; PPV-MF, post-polycythaemia vera-myelofibrosis; RUX, ruxolitinib.
Note: a DIPSS score calculation: 1 point for each of the following criteria: age > 65 years, white cell count ≥ 25 x 109/L, haemoglobin < 10 g/dL, peripheral blood blasts ≥ 1%, constitutional symptoms (weight loss and/or unexplained fever or excessive sweats).

The ERG notes that the eligibility criteria of JAKARTA-2 and SIMPLIFY-2 appear to be different. In SIMPLIFY-2 patients had to either require red blood cell transfusions while on ruxolitinib, or require a

ruxolitinib dose reduction with at least one Grade 3 adverse event of thrombocytopenia, anaemia, or bleeding.

Baseline characteristics for the BAT arms of PERSIST-2 and SIMPLIFY-2, reproduced from Appendix D, Table 14 of the CS, are presented in

Table 18. The CS (Page 54) notes that as no baseline characteristics specific to the JAK-inhibitor exposed population were available for the PERSIST-2 study, only a naïve comparison between fedratinib in JAKARTA-2 and BAT in PERSIST-2 was feasible.

The CS (Page 56) notes that one of the key differences in study inclusion/exclusion criteria was platelet count at baseline; with JAKARTA-2 only including patients with a platelet count of ≥50 x 109/L, PERSIST-2 including patients with ≤100 x 10^9/L and SIMPLIFY-2 not applying a limit. To account for this, the company undertook the naïve comparison of PERSIST-2 and JAKARTA-2 on the subgroup of patients in JAKARTA-2 with a platelet count <100 x 10^9/L (N=64, 66%).

Table 18: Baseline characteristics of patients in BAT arms of PERSIST-2 and SIMPLIFY 2 (reproduced from CS Appendix D, Table 14)

Study	PERSIST-2	SIMPLIFY-2
Treatment	**BAT**	**BAT**
N	72 (ITT)	52 (ITT)
Platelet count x 109/L		
Mean (SD)	NR	126.5 (95.9)
Median (min, max)	NR	NR
MF subtype		
n (%) PMF	43 (59.7)	30 (57.7)
n (%) Post-PV MF	16 (22.2)	12 (23.1)
n (%) Post-ET MF	13 (18.1)	10 (19.2)
Risk status		
n (%) Intermediate-1	13 (18.1)	16 (30.8)
n (%) Intermediate-2	37 (51.4)	28 (53.8)
n (%) High-risk	22 (30.6)	8 (15.4)
JAK2 mutational profile		
n (%) Wild type	NR	12 (23.1)
n (%) Mutant	51 (70.8)	37 (71.2)
n (%) Missing/unknown	NR	3 (5.8)
ECOG PS		
n (%) 0	NR	19 (36.5)
n (%) 1	NR	26 (50.0)
n (%) 2	NR	7 (13.5)
n (%) 3	NR	NA
n (%) Missing	3 (4)	NA
n (%) 0/1	54 (75.0)[c]	45 (86.5)
n (%) 2/3	15 (21)[c]	7 (13.5)
Prior RUX treatment		
n (%) prior RUX	33 (45.8)	52 (100)
n (%) RUX-naïve	39 (54.2)	0
Prior RUX treatment duration		
n (%) < 12 weeks	NR	10 (19.2)
n (%) ≥ 12 weeks	NR	33 (63.5)

Study	PERSIST-2	SIMPLIFY-2
Treatment	**BAT**	**BAT**
n (%) Missing	NR	9 (17.3)
Transfusion dependent, n (%)	14 (19.4)	27 (51.9)
Haemoglobin, g/dL		
Mean (SD)	NR	9.5 (1.6)
n (%) < 10	41 (56.9)	NR
Palpable spleen length, cm		
Median (range)	13 (2, 34)	NR
WBC count > 25 x 109/L, n (%)	14 (19.4)	NR
BMI		
Mean (SD)	NR	26.2 (3.8)
Age (years)		
Mean (SD)	NR	69.4 (7.4)
Median (min, max)	69 (32, 83)	NR
Gender		
n (%) male	39 (54.2)	24 (46.2)
Race		
n (%) White	NR	44 (84.6)
n (%) Asian	NR	NR
n (%) Black/African American	NR	0
n (%) Other	NR	4 (7.7)
n (%) Unknown	NR	8 (15.4)
Mean TSSc (SD) [N]	NR	20.5 (16)

Key: BAT, best available therapy; BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ITT, intention-to-treat; JAK2, Janus kinase 2; MF, myelofibrosis; N, number of patients; NR, not reported; PMF, primary myelofibrosis; post-ET MF, post-essential thrombocythemia myelofibrosis; post-PV MF, post-polycythaemia vera myelofibrosis; RUX, ruxolitinib; SD, standard deviation; SVR, spleen volume reduction; TSS, total symptom score; WBC, white blood cell.
Source: ITC report[33]

Comparator characteristics of studies in the ITC

Given JAKARTA-2 is a single-arm study, the company undertook an ITC to investigate the comparative efficacy and safety of fedratinib versus BAT. BAT in PERSIST- 2 (intervention, Pacritinib) was ruxolitinib (45%), Watch and wait (19%), Hydroxycarbamide (hydroxyurea) (19%), Prednisone (13%), Danazol (5%), Thalidomide (3%), Decitabine (2%), Interferon-alpha (2%). BAT in SIMPLIFY-2

(Momelotinib) was ruxolitinib (89%), Hydroxycarbamide (hydroxyurea) (23%), Corticosteroids (12%). Clinical advice received by the ERG was that, although Decitabine is not approved in the UK for the treatment of myelofibrosis, that the other BAT treatments in both RCTs would be comparable to UK practice. The ERG considers the BAT treatments to be comparable to those of the NICE scope.[2] However, the ERG notes that the NICE Scope does not include ruxolitinib as a comparator for the previously treated MF population.

Outcomes of studies in the ITC

The CS (Page 57) reports that both PERSIST-2 and SIMPLIFY-2 report two efficacy outcomes of interest to the ITC: the proportion of patients achieving ≥35% SVR at 24 weeks from baseline; and the proportion of patients achieving ≥50% TSS reduction at 24 weeks from baseline.

Details of the available data for the proportion of patients with ≥ 35% spleen volume reduction are presented in Table 19 (reproduced from CS, Appendix D Table 15), and details of the available data for the proportion of patients with total symptom score reduction are presented in Table 20 (reproduced from CS, Appendix D Table 16).

Table 19: Available data for the proportion of patients with ≥ 35% spleen volume reduction (reproduced from CS, Appendix D Table 15)

Outcome	PERSIST-2			SIMPLIFY-2	
	Pacritinib 400 mg	Pacritinib 200 mg	BAT (N=72 [ITT])	Momelotinib	BAT (N=52 [ITT])
≥ 35% SVR from baseline to Week 24/EOC 6 for the ITT population	NA*	NA*	NA*	7% (N=104)	6% (N=52)
≥ 35% SVR from baseline to Week 24 for patients with platelet count < 100 x 10^9/L	6%** (N=31)	13%** (N=31)	3%** (N=33)	NR	NR

Key: BAT, best available therapy; EOC, End of Cycle; FEDR, fedratinib; ITT, intention-to-treat; LOCF, last observation carried forward; NA, not applicable; NR, not reported; SVR, spleen volume reduction; wo, without.
Note: *, ITT results for PERSIST-2 include 53% of patients who are JAK-inhibitor naïve; **, results for the subgroup of patients who had received prior ruxolitinib treatment.
Source: ITC report[33]

Table 20: Available data for the proportion of patients with total symptom score reduction (reproduced from CS, Appendix D Table 16)

Outcome	PERSIST-2			SIMPLIFY-2	
	Pacritinib 400 mg (N=75 [ITT wo LOCF])	Pacritinib 200 mg (N=74 [ITT wo LOCF])	BAT (N=72 [ITT wo LOCF])	Momelotinib (N=104 [ITT wo LOCF])	BAT (N=52 [ITT])
≥ 50% reduction in TSS from baseline to 24 weeks for the ITT population	NA*	NA*	NA*	26% (N=103)	6% (N=51)
≥ 50% reduction in TSS from baseline to 24 weeks for the patients with platelet count < 100 x 109/L	10%** (N=31)	32%** (N=31)	15%** (N=33)	NR	NR

Key: BAT, best available therapy; FEDR, fedratinib; ITT, intention-to-treat; LOCF, last observation carried forward; NR, not reported; TSS, total symptom score; wo, without.
Note: *, ITT results for PERSIST-2 include 53% of patients who are JAK-inhibitor naïve; **, results for the subgroup of patients who had received prior ruxolitinib treatment.
Source: ITC report[33]

Spleen or symptom response in the ITC

The CS (Page 107) did not have access to the comparator trial data for BAT and defined the combined endpoint as follows:

- *"The number of BAT patients who reach the endpoint is equal to the maximum number of patients experiencing either SVR or TSS response separately – Referred to henceforth as the Minimum BAT response scenario.*
- *The number of BAT patients who reach the endpoint is equal to the sum of patients experiencing either SVR or TSS response separately – Referred to henceforth as the Maximum BAT response scenario*"

Assuming, for example, five patients are SVR responders and three patients are TSS responders out of 30 patients, then the actual configuration could be one of four tables presented in Appendix 1. The first definition corresponds to the five patients in Appendix 1 Table 1who did not experience "No response" for either SVR or TSS. The second definition corresponds to the eight patients in Appendix 1 Table 4 who did not experience "No response" for either SVR or TSS.

It is not clear to the ERG whether the clinically relevant definition should be with respect to "SVR AND TSS" or "SVR OR TSS". Nevertheless, although the actual configuration is unknown, it is possible to analyse tables such as these after incorporating external information about the correlation between outcomes.[35]

Adverse events of studies in the ITC

Adverse events for the PERSIST-2 and SIMPLIFY-2 RCTs were not reported in the CS.

In PERSIST-2, the most common adverse event leading to discontinuation was thrombocytopenia with pacritinib once daily (4 instances [4%]) and BAT (2 instances [2%]) and anaemia with pacritinib twice daily (3 instances [3%]). The incidence of all hematologic adverse events was similar for patients with baseline platelet count less than 50 × 109/L vs 50 × 109/L or more with pacritinib once daily (27 patients [54%] vs 27 patients [52%]) and pacritinib twice daily (28 patients [60%] vs 33 patients [57%]), but was higher for patients with a baseline platelet count less than 50 × 109/L vs 50 × 109/L or more with BAT (22 patients [52%] vs 21 patients [38%]). Rate of on-study death was lowest with pacritinib twice daily (6 instances [6%]) vs BAT (9 instances [9%]) or pacritinib once daily (14 instances [14%]).[9]

In SIMPLIFY-2, adverse events leading to treatment discontinuation occurred in 22/104 (21%) of the momelotinib group and 1/52 (2%) of the BAT group. Adverse events leading to dose reduction occurred in 17 (16%) of the momelotinib group and 9 (17%) of the BAT group. A similar proportion of patients had Grade 3 or worse anaemia-related AEs in both treatment groups. Deaths due to AEs were reported for six (6%) patients receiving momelotinib and four (8%) patients receiving BAT.[10]

3.4 Critique of the indirect treatment comparison

3.4.1 Methods

The CS (Page 54) reports that an unanchored indirect comparison of fedratinib with BAT as either a matching-adjusted indirect comparison (MAIC) or simulated treatment comparison (STC) was considered. As baseline characteristics were available for SIMPLIFY-2, the CS (Page 54) reports that MAIC and STC analyses, controlling for baseline characteristics identified as being both prognostic and imbalanced between data sources, were explored for comparisons between fedratinib and BAT. However, the CS (Page 54) reports that, given that JAKARTA-2 is a single-arm study, it was not possible to use the patient-level data to identify treatment effect modifiers and no information on treatment effect modifiers for this population was found in the literature. Also, as no baseline characteristics specific to the JAK-inhibitor exposed population were available for the PERSIST-2 study, the CS (Page 54) reports that only a naïve comparison between fedratinib in JAKARTA-2 and BAT in PERSIST-2 was feasible.

The CS (Page 56) notes that there was a key difference in platelet count at baseline across the studies in the ITC; with JAKARTA-2 only including patients with a platelet count of ≥50 x 10^9/L, PERSIST-2 including patients with ≤100 x 10^9/L and SIMPLIFY-2 not applying a limit. To account for this, the CS (Page 56) reports that the naïve comparison of PERSIST-2 and JAKARTA-2 was conducted on the subgroup of patients in JAKARTA-2 with a platelet count <100 x 10^9/L (N=64, 66%).

3.4.2 Results

- The CS (Page 57) reports that in the naïve ITC of JAKARTA-2 and PERSIST-2, comparing fedratinib to BAT in patients with a platelet count <100 x 10^9/L, treatment with fedratinib was associated with a greater proportion of patients achieving ≥35% SVR (■% greater; 95% CI: ■) and a greater proportion of patients achieving ≥50% reduction in TSS (■% greater; 95% CI: ■). Table 21 reproduced from CS Table 17).

Table 21: Summary of the naïve ITC of JAKARTA-2 and PERSIST-2, comparing fedratinib to BAT in patients with a platelet count <100 x 109/ (reproduced from CS Table 17)

Comparison made	Data used to make the comparison	
	JAKARTA-2: fedratinib 400 mg (N=33)[a]	**PERSIST-2: BAT (N=33)[b]**
Proportion of ITT subjects with platelet count <100 x 10^9/L achieving ≥35% SVR from baseline to Week 24/EOC6, n (%)[c]	■ (■%)	1 (3%)
Naïve ITC for ITT subjects with platelet count <100 x 10^9/L achieving ≥35% SVR from baseline to Week 24/EOC6, RD (95% CI)[c]	■% (■)	
Proportion of ITT subjects with platelet count <100 x 10^9/L achieving ≥50% TSS reduction from baseline to Week 24/EOC6, n (%)	■ (■%)	5 (15%)
Naïve ITC for ITT subjects with platelet count <100 x 10^9/L achieving ≥50% TSS reduction from baseline to Week 24/EOC6, RD (95% CI) [c]	■% (■)	

Key: BAT, best available therapy; EOC, end of cycle; ITC, indirect treatment comparison; L, litre; N, total number of subjects; RD, risk difference; SVR, spleen volume reduction; TSS, total symptom score.
Notes: [a], denominator refers to the number of patients from JAKARTA-2 with platelet count of <100 x 10^9/L at baseline; [b], denominator refers to patients from PERSIST-2 that had previously been treated with ruxolitinib; RD calculated by subtracting the proportion of BAT responders from the proportion of fedratinib responders; [c], this row indicates absolute responses and is not an ITC.
Source: ITC report.[33]

The CS (Page 58) reports that adjusted analyses using PERSIST-2 was not possible due to the paucity of publicly available baseline characteristics for the ruxolitinib exposed population. Therefore, the adjusted analyses were conducted using the JAKARTA-2 and SIMPLIFY-2 studies only in the ITT population for JAKARTA-2, presented below. Results were consistent with an ITC conducted using the Sensitivity Cohort from JAKARTA-2.

The CS (Page 58) reports that identification of imbalanced prognostic factors to adjust for in the ITC was performed as follows:

- The variable was identified as imbalanced across the JAKARTA-2 study and the BAT arm of the SIMPLIFY-2 study based on an external haematologist identifying the imbalance as clinically meaningful.
- The variable was identified as being an important prognostic factor based on univariable and multivariable analyses performed with the JAKARTA-2 patient-level data.

Variables fulfilling both criteria for SVR were ECOG PS and transfusion dependence, and variables fulfilling both criteria for TSS reduction were ECOG PS and DIPSS.

The CS ITC results adjusting for prognostic variables are presented in

Table 22 (reproduced from CS Table 18) and

Table 23 (reproduced from CS Table 19).

Table 22: Naïve and adjusted ITC results for SVR (JAKARTA-2 and SIMPLIFY-2) (reproduced from CS Table 18)

Method	Variables included in adjustment[b]	JAKARTA-2 (fedratinib 400 mg; N=97)	SIMPLIFY-2 (BAT; N=52)
Naïve ITC	• NA	30.9% (n=30)	5.8% (n=3)
		RD[c] (95% CI): 25.2% (14, 36.3)	
MAIC	• ECOG PS	[redacted]% (CI: [redacted])[a]	5.8% (n=3)
		RD[c] (95% CI): [redacted]% [redacted])[a]	
STC	• ECOG PS	[redacted]% (CI: [redacted])	5.8% (n=3)
		RD[c] (95% CI): [redacted]% ([redacted])	
MAIC	• ECOG PS • Transfusion dependence	[redacted]% (CI: [redacted])[a]	5.8% (n=3)
		RD[c] (95% CI): [redacted]% [[redacted]][a]	

Key: BAT, best available therapy; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ESS, effective sample size; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; n, number of responders; N, total number of patients; NA, not applicable; RD, risk difference; STC, simulated treatment comparison; SVR, spleen volume reduction.
Note: [a], bootstrap percentile CI (based on 10,000 samples); [b], ESS of JAKARTA-2 population after matching on ECOG PS was 91.7 (94.5% of original sample size) and after matching on ECOG PS and transfusion dependence was 34.4 (35.5% of original sample size); [c] RD calculated by subtracting the proportion of BAT responders from the proportion of fedratinib responders.
Source: ITC report.[33]

Table 23: Naïve and adjusted ITC results for TSS reduction (JAKARTA-2 and SIMPLIFY-2) (reproduced from CS Table 19)

Method	Variables included in adjustment[b]	JAKARTA-2 (400 mg fedratinib; N=97)	SIMPLIFY-2 (BAT; N=51)
Naïve ITC	• NA	[redacted]% (n=[redacted])	5.9% (n=3)
		RD[C] (95% CI): [redacted]% ([redacted])	
MAIC	• ECOG PS • DIPSS	[redacted]% ([redacted])[a]	5.9% (n=3)
		RD (95% CI): [redacted]% ([redacted])	
STC	• ECOG PS • DIPSS	[redacted]% ([redacted])	5.9% (n=3)
		RD (95% CI): [redacted]% ([redacted])	

Key: BAT, best available therapy; CI, confidence interval, DIPSS, Dynamic International Prognostic Scoring System; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ESS, effective sample size; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; n, number of responders; N, total number of subjects; NA, not applicable; RD, risk difference; STC, simulated treatment comparison; TSS, total symptom score.
Note: [a], bootstrap percentile CI (based on 10,000 samples); [b], ESS of JAKARTA-2 population after matching on ECOG PS and DIPSS was 81.6 (84.2% of original sample size); [c], RD calculated by subtracting the proportion of BAT responders from the proportion of fedratinib responders.
Source: ITC report.[33]

The CS (Page 59) reports that as the matching procedure for the MAIC of SVR led to a relatively small effective sample size (ESS) for the JAKARTA-2 population (ESS was 34.4, 35.5% of the original sample size), additional analyses were performed with adjustment for ECOG PS only (in this case the ESS was 91.7).

The CS (Page 59) reports that for the STC of SVR, the adjustment for ECOG PS and transfusion dependence resulted in a logistic regression model that had a standard error of 1,722.4 for the transfusion dependence coefficient compared with a standard error of 0.63 for the ECOG PS coefficient and that the high standard error was likely due to transfusion dependence being a perfect predictor of the outcome and, therefore, the model struggled to converge.

The CS (Page 60) reports that when no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had a 25.2% (95% CI: 14, 36.3) greater proportion of patients with ≥35% SVR compared with BAT. After adjustment for baseline ECOG PS the difference in the proportion of patients with ≥35% SVR compared with BAT increased; fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥35% SVR compared with BAT. After adjustment for baseline ECOG PS and transfusion dependence, fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater

proportion of patients with ≥35% SVR compared with BAT. The CS (Page 60) notes that the results with adjustment for ECOG PS and transfusion dependence should be interpreted with caution given the relatively small effective sample size.

The CS (Page 60) reports that for both the naïve analyses and the MAIC, fedratinib 400 mg led to a greater proportion of patients achieving ≥50% reduction in TSS compared with BAT. When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had an [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥50% TSS reduction compared with BAT. The MAIC, which adjusted for ECOG PS and DIPSS, showed that fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥50% TSS reduction compared with BAT. Similarly, a [redacted]% (95% CI: [redacted]) difference was observed using the STC method.

Adverse events in the indirect treatment comparison

The overall summary of safety in the BAT arms of PERSIST-2 and SIMPLIFY-2 are presented in Table 24 (reproduced from CS Table 20).

Table 24: Summary of treatment emergent AEs reported for JAKARTA-2 and BAT arms of PERSIST-2 and SIMPLIFY-2 (reproduced from CS Table 20)

	JAKARTA-2: fedratinib 400 mg (N=97)	PERSIST-2: BAT (N=98 [Safety population])	SIMPLIFY-2: BAT (N=52)
n (%) of patients with at least one AE	97 (100%)	87 (89%)	46 (89%)
n (%) of patients with at least one Grade 3 or 4 AE	61 (62.9%)	48 (49%)	NR
n (%) of patients with at least one SAE	33 (34.0%)	30 (31%)	12 (23%)
n (%) of patients who discontinued treatment due to AEs	19 (19.6%)	4 (4%)	1 (2%)
n (%) of patients with AEs leading to death	7 (7.2%)	9 (9%)[a]	4 (8%)
n (%) of patients with dose interruption for at least 7 consecutive days	25 (25.8%)	10 (10%)[b]	NR
n (%) of patients with dose reduction	38 (39.2%)	7 (7%)	NR

Key: AE, adverse event; BAT, best available therapy; N, number of patients; NR, not reported; SAE, serious adverse event.
Note: [a], percent is given for N=100; [b], not specified whether the dose interruption was for a least 7 consecutive days.
Source: ITC report.[33]

CS uncertainties in the indirect treatment comparison

The CS (Page 63) reports that identification of treatment effect modifiers was not possible for the ITC analyses as the JAKARTA-2 study is a single-arm trial and there is a paucity of literature on this topic. The variables that could be adjusted for in the ITC analyses were also limited to the reported baseline characteristics from the SIMPLIFY-2 study.

The CS (Page 63) reports that the ITC analyses are also limited by not including adjustment for transfusion dependence, identified by an external haematologist as a baseline characteristic that has a clinically meaningful imbalance between the JAKARTA-2 and SIMPLIFY-2 studies. However, attempts to adjust for transfusion dependence resulted in an ESS of 34.4, therefore estimates using the weights from this adjustment are likely to be unstable. The weights from this adjustment also indicate that a small set of JAKARTA-2 patients were influencing the results.

The CS (Page 64) notes that the JAKARTA-2 and SIMPLIFY-2 studies used different symptom questionnaires to calculate TSS (JAKARTA-2 uses the modified MF-SAF and SIMPLIFY-2 uses Version 2 of the Myeloproliferative Neoplasm Symptom Assessment Form). Therefore, results from the comparison of the percentages of patients achieving ≥50% reduction in TSS should be interpreted with caution.

The CS (Page 64) reports that one of the main limitations to the analyses comparing fedratinib with BAT, where the efficacy of BAT is informed by the PERSIST-2 study, is the unavailability of information for the subgroup of BAT-treated PERSIST-2 patients who had received prior ruxolitinib. Information is not available to understand which treatments patients in this subgroup received; therefore, the composition of BAT is unknown. The baseline characteristics for this subgroup are also not reported meaning it is difficult to conclude how similar patients in JAKARTA-2 and this PERSIST-2 subgroup are. A robust analysis that adjusts for differences in baseline characteristics is also not possible.

The subgroup of JAKARTA-2 patients with a platelet count <100 x 10^9/L was used to compare to the PERSIST-2 evidence. All patients in PERSIST-2 had a platelet count /L. However, even though the information was not available for the subgroup of PERSIST-2 patients who had received prior ruxolitinib, there is likely to still be a disparity in patients with a platelet count <50 x 10^9/L. JAKARTA-2 only included patients with a platelet count ≥50 x 10^9/L, whereas 44% of the ITT BAT-treated PERSIST-2 patients had a platelet count <50 x 10^9/L.

As with the comparison of SIMPLIFY-2 evidence, the PERSIST-2 study used a different symptom assessment form to that used in JAKARTA-2, meaning results from the comparison of the percentages of patients achieving ≥50% reduction in TSS should be interpreted with caution.

The CS (Page 64) reports that expert elicitation was sought to establish whether the composition of BAT in PERSIST-2 and SIMPLIFY-2 studies was representative of how patients would be treated with BAT in the UK. Feedback received during the CS advisory board indicated that PERSIST-2 is not representative of patients receiving ruxolitinib in BAT in the UK as it included patients with platelets $<50 \times 10^9$/L, for which ruxolitinib is not licensed. Specifically, 34 of the 72 patients in the BAT arm of PERSIST-2 had platelets $<50 \times 10^9$/L at baseline. Additionally, many patients with lower platelet count must reduce their ruxolitinib dose as per licensing, so the proportion of ruxolitinib in BAT observed in PERSIST-2 (44%) may be more conservative than what would be seen in UK clinical practice.

3.4.3 ERG critique of the ITC

The ERG has some concerns with the unanchored indirect comparison of fedratinib to BAT. The main concern is whether there are differences in patient populations beyond the attempt to adjust for all relevant prognostic factors and treatment effect modifiers. Furthermore, the ERG has some concerns with some aspects of the methods used to identify prognostic factors and treatment effect modifiers. The ERG suggests that variables included in the International Prognostic Scoring System (IPSS) and the Dynamic International Prognostic Scoring System (DIPSS) should be included in the model irrespective of their statistical significance. Identifying potential prognostic factors based on univariate logistic regression models and forward selection is known to be problematic as is leaving out important prognostic variables because the p-value is non-significant. Continuous variables should be included as continuous variable if possible; categorisation of continuous variables leads to a loss of information and implies that the effect of the covariates changes abruptly. The ERG believes that it is unlikely that the propensity score type model has accounted for all prognostic factors and treatment effect modifiers and that estimates of relative treatment effect may be biased and over-precise.

The CS is transparent in describing many of the limitations associated with the unanchored indirect comparison but not in quantifying the extent of the potential bias and uncertainty associated with the indirect comparison. Furthermore, the ERG has some concerns with some aspects of the methods used to identify prognostic factors and treatment effect modifiers when comparing fedratinib to BAT:

The CS states that risk categorisation was carried out using the International Prognostic Scoring System (IPSS) or the Dynamic International Prognostic Scoring System (DIPSS). Both categorisations include age, white blood cell count, haemoglobin, peripheral blood blasts and constitutional systems. Hence, the ERG suggests that these variables (and not the resulting categories) should be included in any

propensity score type model irrespective of their statistical significance, although SIMPLIFY-2 did not report white blood cell count, peripheral blood blasts and constitutional symptoms. Absence of evidence that a variable is prognostic is not the same as evidence of absence of a variable being prognostic, and external clinical opinion should be used to guide which variable are included in or excluded from the model.

Potential prognostic factors were identified using univariate logistic regression models and multivariable logistic regression models with statistically significant variables identified using forward selection. This approach is known to produce p-values that are too small, regression coefficients that are biased away from zero and standard errors that are too small.[36] On the other hand, leaving out important prognostic variables because the p-value is non-significant is also problematic as discussed above.

The CS considered variables for inclusion in the propensity score type model depending on whether there was an imbalance of ≥ 10 between JAKARTA-2 patients and the pooled BAT-treated patients in PERSIST-2 and SIMPLIFY-2. The ERG does not consider it appropriate to simply pool the evidence on BAT because this ignores potential heterogeneity in the baseline response. Also, in Section 2.9.5.3 of the CS, the company wrote that, "*PERSIST-2 is not representative of patients receiving ruxolitinib in BAT in the UK as it included patients with platelets <50 x 10^9 /L, for which ruxolitinib is not licensed*." Furthermore, the ERG believes that the issue of imbalance is irrelevant; prognostic variables should be included in the model even if they are balanced between treatments (as one should do even in a randomised controlled trial) and variables that are imbalanced are not important if they are not prognostic. Whether consideration is of a response-adjusted model or a treatment-allocation model, the models are non-linear and the measure of "effect" is a non-collapsible measure. Hence, the ERG believes that it is necessary to account for all prognostic factors irrespective of balance in order to estimate the true estimate of effect and the standard error.

The ERG believes that continuous variables should be included in the model as continuous variables when possible. Categorisation of continuous variables leads to a loss of information and implies that the effect of a covariate changes abruptly at the cut-offs rather than with each unit increase in the continuous variable.

In response to clarification question A8,[18] the company stated that it did not match for mean [SD] age, mean haemoglobin, or mean platelet count.

In response to clarification question A8,[18] additional MAICs were performed by the company matching for ECOG PS, DIPSS, transfusion dependence and mean age ([redacted] in JAKARTA-2 and 69.4 [SD7.4] in

the BAT arm of SIMPLIFY-2), but not mean haemoglobin ([redacted] in JAKARTA-2 and 9.5 [SD1.6] in the BAT arm of SIMPLIFY-2), or mean platelet count ([redacted]) in JAKRTA-2 and 126.5 [SD95.9] in SIMPLIFY-2) "*given balance across studies*". After matching, the effective sample size was reduced to [redacted] and the difference in SVR was [redacted] [corrected by the company in subsequent clarification question A4] compared to [redacted]% [95% CI: [redacted]] after adjusting for only ECOG PS and transfusion dependence.

The CS (Section 2.9.5.1) states that, "*Identification of treatment effect modifiers was not possible for these analyses given the JAKARTA-2 study is a single-arm trial and there is a paucity of literature on this topic. The variables that could be adjusted for in these analyses were also limited to the reported baseline characteristics from the SIMPLIFY-2 study.*" The ERG believes that it is unlikely that the propensity score type model has accounted for all prognostic factors and treatment effect modifiers and that estimates of relative treatment effect may be biased and over-precise.

Ultimately, the MAIC of SVR was adjusted for only ECOG PS and transfusion dependence, and the MAIC of TSS for only ECOG and DIPSS. However, there was a much greater proportion of patients who were transfusion dependent in SIMPLIFY-2. The company suggest that SVR results after adjustment for ECOG PS and transfusion dependence should be treated with caution because of the relatively small effective sample size. The ERG suggests that the issues are whether the population defined by the SIMPLIFY-2 study is representative of the target population and whether all relevant prognostic factors and treatment effect modifiers have been accounted for; the small effective sample size is accounted for through the estimated confidence interval.

Estimates of relative treatment effect are presented on the absolute risk scale rather than the odds ratio (or log-odds ratio scale) which is assumed to be the additive scale on which relative treatment effects are estimated. In response to clarification question A9,[18] the company wrote:

"*For the MAIC analyses, a risk difference was calculated by:*

1. *Simulating the SIMPLIFY-2 BAT data based on the number of reported responders and non-responders*
2. *Combining the simulated comparator data with the JAKARTA-2 IPD*
3. *Fitting a binomial model with logit link to the combined data that has treatment as a covariate and includes the weights (simulated comparator subjects were assigned a weight of 1)*
4. *Finally, the proportion of comparator responders predicted from the model was subtracted from the proportion of fedratinib responders, also predicted from the model*"

The ERG notes that the risk difference is unlikely to be generalisable and that the primary purpose of back-transforming to treatment-specific absolute risks is for use in the economic model. Consequently, (CS, Table 44) the ERG suggests that the way the results have been used is inappropriate. The analyses generates a relative treatment effect of fedratinib versus BAT, in this case in the ITT population. The absolute response to fedratinib should be computed by adding the relative effect to the BAT response on the additive scale and not by subtracting the difference in absolute responses from the fedratininb response as the company has done.

In response to clarification question A12,[18] the company wrote that none of the available studies (summarised in Appendix L.5.1) included "*OS Kaplan-Meier* [survival functions] *and report sufficient data on baseline characteristics*" so that an unanchored indirect comparison of overall survival could not be conducted. Nevertheless, the company claims in Section B2.12 of the CS that fedratinib delivers a survival gain. The ERG notes that an effect on overall survival is unproven.

3.5 Additional work on clinical effectiveness undertaken by the ERG

No additional work was undertaken by the ERG.

3.6 Conclusions of the clinical effectiveness section

The ERG considers that the company's search strategy is sufficiently comprehensive to retrieve important citations relating to clinical effectiveness and safety of fedratinib for splenomegaly and symptoms in myelofibrosis, although some limitations are noted.

The target population defined by the company is patients with MF who have previously been treated with ruxolitinib. This is one of the populations in the NICE scope[2], but the CS does not consider patients who have not previously been treated with ruxolitinib which was also within the NICE scope. The key clinical evidence submitted by the company is derived from a single-arm, open-label, non-randomised, phase 2, multicentre study (JAKARTA-2); of fedratinib in myelofibrosis patients who were currently or previously treated with ruxolitinib for at least 28 days and who had symptomatic intermediate-1 risk, intermediate-2 or high-risk disease.[11] Patients had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib and therefore were not currently treated with ruxolitinib nor were they previously treated for at least 28 days.

The ERG has some concerns with the design of the JAKARTA-2 study in that: the estimates of treatment effects from Phase 2 studies generally over-estimate true treatment effects, populations defined by Phase 2 inclusion/exclusion criteria are often less diverse than in Phase 3 studies, the lack of a concurrent control means that the study is likely to suffer from the regression to the mean, and

although the primary outcome measure may be considered clinically relevant, dichotomising a continuous variable is statistically inefficient.

In the ITT population (N=97) of the JAKARTA-2 study, 31% (95%CI 22 to 41) of patients achieved the primary outcome of spleen response rate defined as ≥35% SVR at EOC6. In the ITT population (N=97), 25% (95%CI 17 to 35) of patients achieved the key secondary outcome of symptom response rate defined as ≥50% reduction in TSS at EOC6. The 12-month overall survival rate in the ITT population (N=97) was [redacted].

In the ITT population (N=97) of the JAKARTA-2 study, all patients had at least one treatment-emergent adverse event (TEAE) of any grade. Grade 3 or 4 TEAEs were reported by [redacted] (63%) patients, including transfusion dependency in [redacted] ([redacted]%) patients. Treatment-emergent serious adverse events (SAEs) were reported by [redacted] (34%) patients. Seven (7%) patients had a TEAE that led to death during treatment or follow-up. TEAEs leading to treatment discontinuation occurred in [redacted] (20%) patients and TEAEs leading to dose modification occurred in [redacted] ([redacted]%) patients.

In the ITC, MAIC and STC analyses, controlling for baseline characteristics identified as being both prognostic and imbalanced between data sources, were explored for comparisons between fedratinib and BAT using baseline characteristics from one study (SIMPLIFY-2). As no baseline characteristics specific to the JAK-inhibitor exposed population were available for the other study in the ITC (PERSIST-2), only a naïve comparison between fedratinib in JAKARTA-2 and BAT in PERSIST-2 was feasible. In the naïve ITC of JAKARTA-2 and PERSIST-2, comparing fedratinib to BAT in patients with a platelet count <100 x 10^9/L, treatment with fedratinib was associated with a greater proportion of patients achieving ≥35% SVR ([redacted]% greater; 95% CI: [redacted]) and a greater proportion of patients achieving ≥50% reduction in TSS ([redacted]% greater; 95% CI: [redacted]). When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400 mg had a 25.2% (95% CI: 14, 36.3) greater proportion of patients with ≥35% SVR compared with BAT. After adjustment for baseline ECOG PS the difference in the proportion of patients with ≥35% SVR compared with BAT increased; fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥35% SVR compared with BAT. After adjustment for baseline ECOG PS and transfusion dependence, fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥35% SVR compared with BAT. When no adjustment was made for differences in prognostic factors or treatment effect modifiers, fedratinib 400mg had an [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥50% TSS reduction compared with BAT. The MAIC, which adjusted for ECOG PS and DIPSS, showed that fedratinib 400 mg had a [redacted]% (95% CI: [redacted]) greater proportion of patients with ≥50% TSS reduction compared with BAT. Similarly, a [redacted]% (95% CI: [redacted]) difference was observed using the STC method.

The ERG has some concerns with the indirect comparison of fedratinib to BAT. The main concern is differences in patient population beyond adjusting for all relevant prognostic factors and treatment effect modifiers dealt with the fact that (CS Section 2.3.1.1), "*Of the patients enrolled and treated in JAKARTA-2, the majority (■%) were resistant to ruxolitinib, ■ (■%) were intolerant to ruxolitinib and one patient (■%) was neither resistant nor intolerant and was categorised as 'other: lack of efficacy'*." In contrast, 54.2% of patients in PERSIST-2 where ruxolitinib naïve and patients in SIMPLIFY-2 could be current treated with ruxolitinib. The ruxolitinib-related difference in populations may reflect a difference in populations that is not possible to adjust for by matching according to baseline characteristics that are prognostic factors and treatment effect modifiers alone.

The ERG notes that the CS is transparent in describing many of the limitations associated with the unanchored indirect comparison but not in quantifying the extent of the potential bias and uncertainty associated with the indirect comparison. Furthermore, the ERG has some concerns with some aspects of the methods used to identify prognostic factors and treatment effect modifiers when comparing fedratinib to BAT. Estimates of relative treatment effect in the CS are presented on the absolute risk scale rather than the odds ratio (or log-odds ratio scale) which is assumed to be the additive scale on which relative treatment effects are estimated. The ERG believes that continuous variables should be included in the model as continuous variables when possible.

4 COST EFFECTIVENESS

This chapter presents a summary and critique of the company's health economic analyses of fedratinib for the treatment of adult patients with MF previously treated with ruxolitinib.

Section 4.1 describes and critiques the company's review of existing economic evaluations. Section 4.2 describes the company's economic model and summarises the company's results. Sections 4.3 and 4.4 present the ERG's critical appraisal of the company's model and the ERG's exploratory analyses. A discussion of the company's economic analysis is provided in Section 4.5. Section 5 presents results from the ERG's exploratory analysis.

4.1 ERG's comment on company's review of cost-effectiveness evidence

4.1.1 Summary and critique of the company's search strategy

The company undertook systematic literature searches to identify i) economic evaluations for any intervention in adult patients with MF with intermediate and high risk (CS Appendix G) ii) utility studies (CS Appendix H) iii) cost and resource use studies (CS Appendix I). The search strategies were not presented at the time of the company submission. The company responded to clarification question 2 on literature searching[18] by providing search strategies to allow the full assessment of the company searches by the ERG.

Initial searches were run on the 18th December 2018 and were updated on the 28th October 2019. The searches covered MEDLINE (via Embase.com), EMBASE (via Embase.com), MEDLINE in process (via PubMed), Econlit (via EBSCO), NHS Economic Evaluations Database and HTA database (via Wiley).

Conference abstracts published between 2017 and 2019 were also searched for the following conferences; International Society for Pharmacoeconomics and Outcomes Research (ISPOR), American Society of Haematology (ASH), British Society for Haematology (BSH) and European Haematology Association (EHA). Key international HTA websites were searched for HTA evaluations and models: (NICE, Scottish Medicines Consortium [SMC], Canadian Agency for Drugs and Technology in Health [CADTH], French National Authority for Health [HAS] and the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA).

The company states (CS, appendix G1) that "*the search strategy was adapted from the economic terms recommended by the Scottish Intercollegiate Guidelines Network (SIGN) (which in turn is recommended by NICE)*." The economic modelling filter used and reported in the company's search strategies differ considerably from SIGN's economic studies search filter on their most up-to-date web

site (https://www.sign.ac.uk/what-we-do/methodology/search-filters/). By contrast to the modelling filter used by the company, SIGN have used and adapted the economic studies filter by the NHS CRD (University of York). There is no clear recommendation from NICE or other HTA bodies, that SIGN's search filter should be used for searching economic studies.

The ERG have only received the company's update searches for all three searches (economic evaluations, utility studies and cost and resource use studies), from September 2019 until February 2020 in all MEDLINE and Embase searches (CS company response[18] Economic SLR, Table 1, page 3). The justification for not providing the initial searches in 2018 are unclear. The ERG is therefore unable to confirm that all eligible studies have been consistently retrieved by the company's SLR economic search.

4.1.2 Summary of company's review findings

A total of nine studies were included (15 publications) in the company's economic systematic review. The company's searches did not identify any economic analyses for either fedratinib or other interventions in people previously treated with ruxolitinib. All studies compared ruxolitinib versus BAT or Placebo. Further details of included and excluded studies are presented in CS Appendix G. Overall, the ERG considers the review to be adequate.

4.2 Summary of the company's submitted economic evaluation

The description of the economic model submitted by the company presented in this ERG report is largely based on information contained within the CS.[1] In the instances where the description of the model's logic and input parameters contained within the CS is brief, unclear and/or inaccurate, the model is used as the basis for this description.[37]

4.2.1 Population

The company developed a de novo economic model[37] to evaluate the cost-effectiveness of fedratinib versus best available therapy (BAT) in the UK in patients with myelofibrosis (MF) previously treated with ruxolitinib.

The CS base-case[1] focuses on the subset of patients with intermediate-2/high risk MF from JAKARTA-2.[11] While the results are not presented within the submission,[1] the economic model submitted by the company[37] includes the functionality to assess the cost-effectiveness of fedratinib in the overall population of JAKARTA-2 (including intermediate-1 with symptoms), and in a subgroup of patients with a baseline platelet count < 100 x 10^9/L (in this scenario, only response rate, platelet count

distribution and proportion of ruxolitinib use is changed). However, results of these groups are not presented in the CS[1].

Baseline characteristics of the intermediate-2/high risk MF patients are: mean age [redacted] years, gender distribution [redacted] male, mean weight [redacted] kg, mean body surface area (BSA) [redacted] m^2, and platelet count distribution([redacted]% ≥ 100,000 x 10^9/L), and are taken directly from JAKARTA-2.

4.2.2 Approach and structure of the economic model

The economic model[37] submitted by the company is built in Microsoft Excel® and uses an individual patient-level simulation approach, described in the CS[1] as a discrete event simulation (DES) approach.

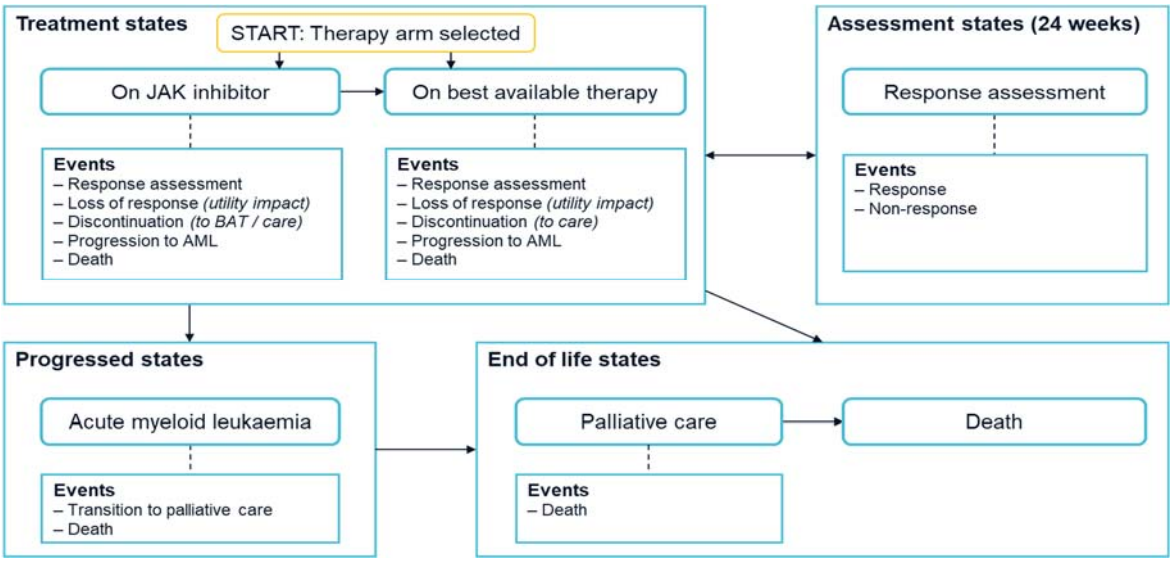
Compared with cohort models, where outcomes are modelled for a typical cohort of patients, in an individual patient-level simulation model, outcomes in the CS are estimated for each individual patient entering the simulation, one at a time, using a stochastic (random) process. Results in the CS are generated assuming 1,000 patients enter the simulation.

The use of an individual patient-level simulation approach (referred to as a DES in the CS) was justified to: (a) mirror the approach used in TA386[13] for the assessment of ruxolitinib in patients with MF, (b) enable memory to be implemented in the model and (c) provide flexibility (as listed in TA386[13]) in dealing with transitions, response assessment at 24 weeks, capturing the progressing nature of MF and exploring alternative structural assumptions.

The company model structure is depicted in

Figure 9 (reproduced from Figure 15 in CS,[1] page 89) and comprises five mutually exclusive health states named; (a) on JAK inhibitor, (b) on BAT, (c) AML (acute myeloid leukaemia), (d) Palliative care and (e) Death (absorbing health state). Patients alive and on treatment at 24 weeks (in the JAK inhibitor or BAT health states) are then categorised onto responders and non-responders, where response is assessed according to either spleen response only (defined as ≥ 35% spleen volume reduction from baseline at 24 weeks), symptom response only (defined as ≥ 50% TSS reduction from baseline at 24 weeks) or a combination of both spleen response and symptom response (used in the base-case). Response is used in the economic model to: (a) assign time to treatment discontinuation (TTD) in patients initiated on fedratinib (subsequently moving to BAT) according to response and (b) assign utility values (and duration of response). Response in the BAT health state is only used to assign utility values (and duration or response).

Figure 9: Structure of the economic model (reproduction of Figure 15 in CS, page 89)



Key: AML, acute myeloid leukaemia; BAT, best available therapy; JAK, Janus kinase.

Patients initiated on fedratinib enter the model in the "on JAK inhibitor" health state and are separated into 4 groups:

- **Group 1**: Early death; defined as patients who die prior to the response assessment at 24 weeks.
    - Patients in this group move directly to the death state according to the OS survival function for patients receiving treatment
- **Group 2**: Early discontinuation; defined as patients alive at 24 weeks, but discontinue treatment prior to the response assessment at 24 weeks,
    - Patients in these group are assumed to move to the BAT health state at the time of discontinuation.
- **Group 3**: Responders at 24 weeks; defined as patients alive and on treatment at 24 weeks, entering the response assessment state and achieving a response according to the chosen definition of response,
    - Patients in this group remain in the JAK inhibitor health state (fedratinib) until they experience any one of the following events (a) treatment discontinuation at which time they move to the BAT health state, (b) transformation to AML at which time move to the AML health state or (c) death
    - Patients with an initial response at 24 weeks, can lose their response whilst on treatment.
- **Group 4**: Non-responders at 24 weeks; defined as patients alive and on treatment at 24 weeks, entering the response assessment state and not achieving a response according to the chosen definition of response
    - Patients in this group follow the same pathway as responders (Group 3).

Patients are initiated on BAT from model initiation or, following discontinuation of fedratinib, enter the "on BAT" health state and are split into 4 groups as per the "on JAK inhibitor" health state. However, compared with the "on JAK inhibitor" health state, no TTD is explicitly modelled for these patients, and therefore patients in the "on BAT" health state remain on treatment until they: (a) transform to AML at which time they move to the AML health state, (b) enter the palliative care health state at the end of life [last 8 weeks] or (c) death.

It should be noted that whilst patients are separated onto group categories, OS for each group is not modelled separately. Therefore, this is not a response-based model. This is further discussed in Section 4.3.4.2.

4.2.3 Interventions and comparators

Intervention

The intervention is fedratinib 400 mg, taken orally once daily, in line with JAKARTA-2[11] and the expected marketing authorisation (CS,[1] page 98).[15]

The CS base-case[1] assumes fedratinib will be given as per the JAKARTA-2 trial protocol (e.g., no stopping rule).[11] A scenario analysis is presented in the CS, where patients who do not respond at 24 weeks (based on the continuation rule for ruxolitinib in the British Committee for Standards in Haematology [BCSH] guideline for the diagnosis and management of myelofibrosis, and used in TA386[13]), stop incurring costs after week 24. A critique of this scenario is presented in Section 4.3.4.10.

Comparator

The comparator in the CS economic model is BAT, consisting of a basket of multiple therapies, including ruxolitinib, based on the BAT composition in SIMPLIFY-2[10] (see Section 4.2.9).

While the CS[1] is in patients previously treated with ruxolitinib, the company considers the majority of BAT to be ruxolitinib (≈89%) as per SIMPLIFY-2.[10] This is justified in the CS[1] following discussion with the company's clinical experts because: (a) the lack of alternative effective targeted therapeutic options available for patients with MF following ruxolitinib failure, and therefore patients are expected to continue on ruxolitinib while achieving a suboptimal response and (b) the possibility for patients to experience rebound and lose symptom control following ruxolitinib discontinuation (CS,[1] page 98).

The use of a basket of therapies (BAT) as a comparator is justified by the company[1] to reflect treatments typically considered in trials in MF and previous economic evaluations for ruxolitinib.[13] The company[1]

further consider ruxolitinib alone not to be a relevant comparator, but included as part of BAT, stating that this is so that costs and efficacy inputs are aligned (CS,[1] page 98-99).

4.2.4 Perspective, time horizon and discounting

While there is no clear statement regarding the perspective, the company states that the reference case has been adhered to (CS,[1] Table 1, page 9). Therefore, costs in the CS are considered from the perspective of the NHS and Personal Social Services (PSS). Both costs and QALYs in the CS are discounted at 3.5% per annum, with LYs undiscounted. Cycle length and half cycle correction were not required in the CS, as the model uses a time to event simulation approach. A lifetime horizon is used by the CS in the base-case (assumed to be 30 years) with shorter time horizon explored in scenario analysis.

4.2.5 Treatment effectiveness and extrapolation in the base-case

Key efficacy inputs include:

- Overall survival (OS)
- Response rate assessed at 24 weeks
- Duration of response (DoR)

These are described in turn below.

4.2.5.1 Overall survival

The description for the modelling of OS focuses on the company's base-case. OS is modelled directly (e.g., OS is not linked to the time in previous health state or response rate). The ERG notes that the company explore an additional scenario analysis where OS is estimated based on a surrogacy assumption. This analysis relies on a number of assumptions and is only presented by the company as a scenario analysis,[1] rather than an alternative base-case. Consequently, no description is presented in this section and the ERG refers the reader to Appendix L.7 of the CS[37] for further details. A critique of this scenario is however presented in Section 4.3.4.9.

OS for fedratinib and comparator is taken from two separate sources and is not adjusted for differences in patient characteristics, so the final comparison is a naïve indirect comparison.

The process for survival model selection was similar for patients initiated on the fedratinib or comparator arm and is therefore described here. The company states that prior to an advisory board,[27, 37] seven clinicians were asked to provide expectations for survival for patients initiated post-ruxolitinib

and on fedratinib,[33] with the mean value subsequently presented to clinicians at the advisory board, alongside information on the predicted survival and statistical goodness-of-fit (AIC and BIC).

- ***Overall survival in patients initiated on fedratinib***

The company fitted six parametric distributions (exponential, Weibull, Gompertz, Log-normal, Log-logistic, and Generalised Gamma) to the OS data from JAKARTA-2[11] in the subset of patients with intermediate-2 high risk (CS,[1] Figure 21, page 129).

The company identified the Gompertz function (Figure 10) as being the most appropriate function to use to represent OS based on clinical plausibility following discussion at the CS advisory board.[27, 37]

Figure 10: Parametric survival functions fitted to overall survival data in the JAKARTA-2 intermediate-2 and high-risk population (Reproduction of Figure 21, CS, page 129)

Key: KM, Kaplan–Meier.

Scenario analysis are presented for a limited number of alternative parametric distributions, reported in the CS appendices.[37]

While a Gompertz distribution is used for fedratinib to extrapolate OS, the company[1] notes that the OS survival function for fedratinib crossed the OS survival function for BAT, which was deemed implausible by clinical experts consulted during the CS advisory board. Consequently, the company[1]

assumed that OS for patients on fedratinib was equal to the OS for patients on BAT following crossing of the survival functions.

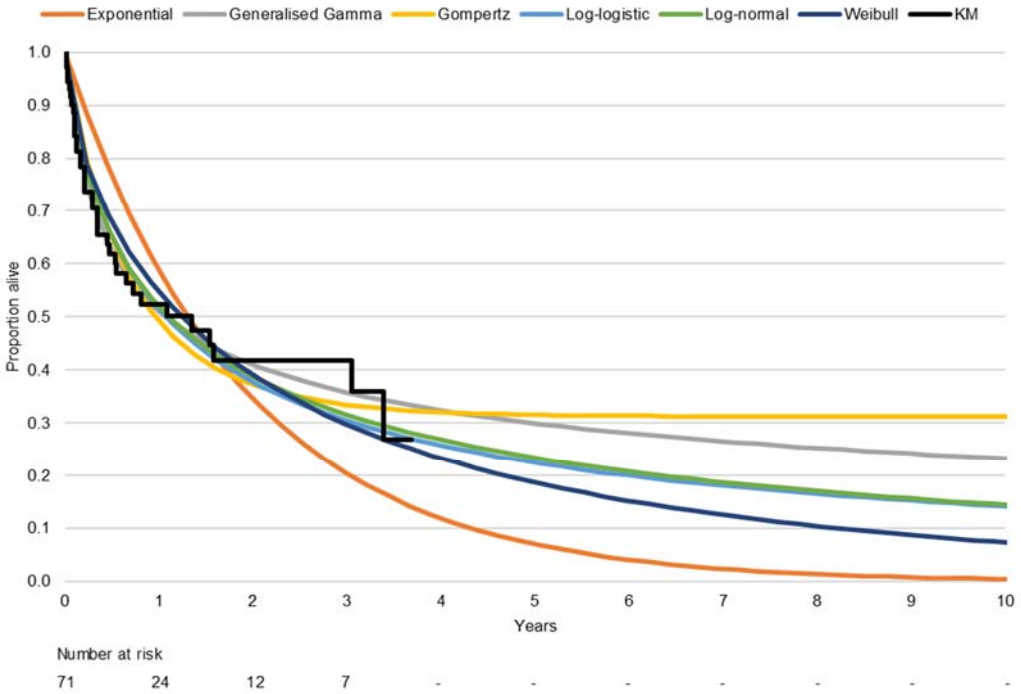
- ***Overall survival in patients initiated on BAT***

As JAKARTA-2[11] is a single-arm trial, the company[1] conducted a systematic review of the literature to identify external sources for OS after discontinuation of ruxolitinib for the BAT arm. Searches were initially conducted in August 2018, and updated in February 2019. A total of 13 studies were included in the OS SLR, with evidence from 4 studies (Schain et al, 2019;[38] TA386,[13] Kuykendall, 2017[39] and Palandri et al, 2019[40]) subsequently included in the economic model. These were selected based on (a) relevance and (b) availability of the KM survival function (CS,[1] page 123).

OS taken from the Schain et al (2019)[38] study is selected in the base-case following discussion with clinical experts at the CS advisory board.[27, 37] Briefly, this study[38] reports survival data for 71 patients in Norway and Sweden who discontinued ruxolitinib, with the most common treatment received following ruxolitinib being glucocorticoids (65.9%) followed by hydroxyurea (32.4%).

The company fitted six parametric distributions (exponential, Weibull, Gompertz, Log-normal, Log-logistic, and Generalised Gamma) to the OS data from Schain et al (2019)[38] (CS,[1] Figure 18, page 126), with the Weibull distribution (Figure 11) selected in the base-case based on both clinical plausibility, and statistical goodness-of-fit. Scenario analysis are presented for a limited number of alternative parametric distributions and sources; reported in the CS appendices.[37]

Figure 11: Parametric survival functions fitted to overall survival data in Schain et al., 2019 (Reproduction of Figure 18, CS, page 126)



Key: KM, Kaplan–Meier.

4.2.5.2 Response rate at 24 weeks

In the CS base-case,[1] response rate at 24 weeks is based on both spleen (volume) and symptom response following the International Working Group Myeloproliferative Neoplasms Research and Treatment guidelines,[41] as well as clinical opinion. Spleen response (volume) alone or symptom response alone are explored in scenario analysis.

Response rate for patients initiated on fedratinib in the CS is based on the number of responders observed in the subset of JAKARTA-2[11] ([redacted]; n = [redacted]) with intermediate-2/high risk. The response rate for patients initiated on BAT ([redacted]%) is derived in the CS from: (a) the response rate above in patients initiated on fedratinib,[11] as specified above (b) the average treatment effect ([redacted]%; differences in proportion calculated between fedratinib and BAT) estimated under an optimistic scenario (maximum response [redacted]%) and a pessimistic scenario (minimum response [redacted]%) using the ITT population (N=97, including intermediate-1 patients) calculated from the MAIC[1] between JAKARTA-2[11] and SIMPLIFY-2,[10] adjusted for ECOG PS and DIPSS only.

Scenario analyses of response rate in the CS are conducted with response assessed using spleen only criteria, symptom only criteria, STC criteria, data from PERSIST[9] and data from a sensitivity cohort. The results for these scenarios are available in Appendix L of the CS.[37]

4.2.5.3 Duration of Response

Duration of response is applied in the CS to responders, from weeks 24 to account for the fact that patients may lose their response prior to discontinuing treatment.

Duration of response is calculated in the CS in patients with a spleen response only, and is based on the duration of spleen response, as per the definition used in JAKARTA-2[11] and in the absence of information on symptoms.

Six parametric distributions (exponential, Weibull, Gompertz, Log-normal, Log-logistic, and Generalised Gamma) are fitted to the DoR data from JAKARTA-2[11] (CS,[1] Figure 16, page 117) in patients with intermediate-2/ high risk (n = [redacted]). The log-normal distribution (Figure 12) was selected based on statistical goodness-of-fit only (CS,[1] page 116). It should be noted that there was only 2 events.

4.2.5.4 Time to treatment discontinuation

TTD is not modelled in the CS for patients entering the BAT health state, and therefore patients in the "on BAT" health state remain on treatment until they: (a) transform to AML, (b) enter the palliative care health state at the end of life [which lasts 8 weeks] or (c) die. The company[1] justify this assumption to reflect the lack of an alternative effective targeted therapy (CS,[1] page 132).

Figure 12: Parametric survival functions fitted to duration of response data in JAKARTA-2 Int-2/High-risk population (Reproduction of Figure 16, page 117)

Key: KM, Kaplan–Meier.

In contrast, patients in the "on JAK inhibitor" health state (only patients initiated on fedratinib) are allowed to discontinue treatment, and move to BAT (excluding ruxolitinib), AML, palliative care and death.

TTD assumed according to the assigned group is described here

- **Group 1**: Early death; defined as patients who die prior to the response assessment at 24 weeks.
    - TTD is based on OS (as these patients die before 24 weeks)
- **Group 2**: Early discontinuation; defined as patients alive at 24 weeks, but discontinue treatment prior to the response assessment at 24 weeks [% (n =)],
    - Time to discontinuation (in weeks) is drawn randomly from a uniform distribution with limits 0 and 24.
- **Group 3**: Responders at 24 weeks; defined as patients alive and on treatment at 24 weeks, entering the response assessment state and achieving a response according to the chosen definition of response,
    - The company fitted six parametric distributions (exponential, Weibull, Gompertz, Log-normal, Log-logistic, and Generalised Gamma) to the TTD data (n =) from JAKARTA-2[11] (CS,[1] Figure 24, page 135) for responders from Week 24 onward from the subset of patients with intermediate-2/high risk MF (Figure 13). The exponential distribution is selected for its clinical plausibility according to the company.

Group 4: Non-responders at 24 weeks; defined as patients alive and on treatment at 24 weeks, entering the response assessment state and not achieving a response according to the chosen definition of response

- The company fitted six parametric distributions (exponential, Weibull, Gompertz, Log-normal, Log-logistic, and Generalised Gamma) to the TTD data (n = [redacted]) from JAKARTA-2[11] (CS,[1] Figure 23, page 134) for non-responders from Week 24 onward (with response defined according to spleen only) in the subset of patients with intermediate-2/high risk. The company[1] justify using TTD estimated in patients with no spleen response (n = [redacted]) compared with patients with no spleen or symptom response (as used in the CS base-case because of issues with fitting parametric distributions given the small sample size (n = [redacted]). The Gompertz distribution is selected in the base-case (Figure 14). The company recognise that this distribution was not associated with the best statistical fit. The following reasons were stated as justification for using the Gompertz distribution by the company[1] (a) expectation that time on treatment is limited for non-responders, (b) consistency with TTD for responders and (c) alternative distributions were associated with a plateau and therefore were considered clinically inappropriate.

Figure 13: Parametric survival function fitted to responder time-to-discontinuation data post-week 24 in JAKARTA-2 Int-2/High-risk patients (Reproduction of Figure 24, CS, page 135)

[redacted]

Key: KM, Kaplan–Meier

Figure 14: Parametric survival functions fitted to non-responder time-to-discontinuation data post-week 24 in JAKARTA-2 Int-2/High-risk patients (Reproduction of Figure 23, CS, page 134)

Key: KM, Kaplan–Meier.

4.2.6 Transition to AML, death following AML transformation and palliative care

4.2.6.1 AML transformation

The same rate of AML transformation is used in the CS base-case in patients initiated on fedratinib or BAT. The company[1] justify this by stating it is unclear if treatment influences the rate of progression to AML (CS,[1] Table 29, page 93).

While the CS[1] does not provide the source or the rate assumed, this can be inferred from the CS economic model[37] directly. While it is stated in the model[37] that the rate is informed by long-term ruxolitinib data,[13] if a constant rate between treatments is selected, the ERG notes that the weekly rate of AML transformation used in the CS economic model ([redacted]) is actually calculated from the AML transformation rate from JAKARTA-2[11] in patients treated with fedratinib from the ITT population (n = [redacted], based on a mean exposure time of [redacted] weeks).

4.2.6.2 Death following AML

The time to death following AML transformation in the CS is taken from Mesa et al (2005)[42] with the generalised gamma fitted to the data. In the CS base-case,[1] the time to death is re-estimated in patients who transform to AML to account for their sample time to death due to other causes. Details on the approach to re-estimate survival is provided in Appendix L.8. of the CS[37]

4.2.6.3 Transition to palliative care

Patients in the BAT health state in the CS (initiated from model start or entering this health state following fedratinib) are assumed to spend the last 8 weeks of life, or time to death if this is estimated to be shorter, in the palliative care health state. This is applied in the CS model retrospectively as OS is modelled directly. For instance, for two hypothetical patients;

- a patient on BAT, with a sampled time to death of 100 weeks. This patient will spend 92 weeks in the BAT health state and 8 weeks in the palliative care health state.
- a patient on BAT, with a sampled time to death of 3 weeks. This patient will spend 0 weeks on BAT and 3 weeks in the palliative care health state.

The ERG notes that patients with AML transformation also enter this health state. Therefore, those with AML who are predicted to live less than 8 weeks would enter the palliative health care directly. Patients in the "JAK inhibitor" health state cannot enter this health state, thus, patients initiated on fedratinib from model initiation can only enter this health state following treatment discontinuation, after they enter the BAT health state.

4.2.7 Adverse events

Only non-haematological adverse events (AEs) grade ≥ 3 included in TA386[13] were considered in the CS. The frequency of non-haematological adverse events Grade ≥3 assumed in the economic model are taken from JAKARTA-2[11] for patients receiving fedratinib, SIMPLIFY-2[10] for patients receiving BAT from model initiation and COMFORT-II[13] for patients receiving BAT following fedratinib discontinuation. The frequency of Grade ≥3 adverse events assumed in the economic model is summarised in Table 25 and is used to estimate management costs only in the base-case.

4.2.8 Health-related quality of life

Patients in JAKARTA-2[11] completed a health-related quality of life questionnaire using the Myelofibrosis Symptom Assessment Form Version 2.0 (MF-SAF V2.0) diary,[43] Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF) questionnaire,[43] the European Organisation for Research and Treatment of Cancer Quality of Life 30 Questionnaire Version 3.0 (EORTC QLQ-C30 V3.0),[43] and the Patient Global Impression of Change (PGIC).[43]

The MF-SAF V2.0 and EORTC QLQ-C30 V3.0 were used in the CS to derive utility values using the MF-8D[44] and EORTC-8D[45] respectively. The MF-SAF V2.0 was completed by patients daily through the first six cycles, via an electronic diary. The EORTC QLQ-C30 V3.0 was completed by patients on Day 1 of each treatment cycle up to Cycle 6, end of Cycle 6, Day 1 of Cycle 13, end of treatment, and at a 30-day follow-up visit.

Table 25: Frequency of Grade ≥3 adverse events assumed in the economic model

	JAKARTA-2[11] (mean exposure: [redacted] years)		**SIMPLIFY-2[10]** (mean exposure: 0.462 years)		**COMFORT-II[13]** (mean exposure: 0.912 years[a])	
Adverse event	**N**	**N**	**n**	**N**	**n**	**N**
Abdominal pain	[redacted]	81	3	52	3	73
Arthralgia	[redacted]	81	0	52	0	73
Asthenia	[redacted]	81	1	52	1	73
Back pain	[redacted]	81	0	52	0	73
Bronchitis	[redacted]	81	0	52	1	73
Cough	[redacted]	81	0	52	1	73
Diarrhoea	[redacted]	81	1	52	0	73
Dyspnoea	[redacted]	81	1	52	3	73
Fatigue	[redacted]	81	1	52	0	73
Headache	[redacted]	81	1	52	0	73
Nausea	[redacted]	81	1	52	0	73
Oedema peripheral	[redacted]	81	0	52	1	73
Pain in extremity	[redacted]	81	0	52	0	73
Pyrexia	[redacted]	81	0	52	0	73
Weight increased	[redacted]	81	0	52	0	73

[a] Taken from the economic model

The MF-8D[46] used in the CS[1] base-case is a condition-specific preference-based measure which was developed and subsequently used in TA386[13] for patients with MF. Scenario analyses were conducted in the CS in which the EORTC-8D[45] and alternative values from the literature base were used. EQ-5D mapping was not presented in the CS because of concern with the ability of the EQ-5D to detect clinically meaningful changes in HRQoL in patients with MF.[44, 47]

Utility values are assigned in the CS according to response status (and duration of response) and applied as change from baseline at 4 weeks, depending on whether patients are classified as responder or not (further details are provided in the CS[1]). Health state utility values are estimated in the CS from a mixed-effect model, to account for multiple observations per patient, with baseline utility value, response status at the end of Cycle 6 and gender as covariates. Statistical models are presented in the CS for different definitions of response (spleen only response, symptom only response, both spleen response or

symptom response) with the latter used in the base-case (to align with the response definition in the base-case).

The utility value for the AML health state in the CS was taken from Pan et al (2010)[48] in patients with secondary AML, with the same utility value assumed for the palliative care health state. Utility values are age-adjusted in the CS using Ara et al (2010).[49] Utility values used in the CS economic model[1, 37] are summarised in Table 26.

Table 26: Utility values used in the CS (Adaptation of Table 64, in CS, page 148-149)

State	Assignment	Utility value: mean (standard error)	95% confidence interval
Baseline utility	Baseline utility use for first 4 weeks after patient first receives treatment	Female: [redacted] ([redacted])	[redacted]
		Male: [redacted]	[redacted]
Treatment: response	Change from baseline at 4 weeks if patient is classified as a responder	Female: [redacted]	[redacted]
		Male: [redacted]	[redacted]
Treatment: non-response	Change from baseline at 4 weeks if patient is classified as a non-responder	Female: [redacted]	[redacted]
		Male: [redacted][a]	[redacted]
Treatment: loss of response	Change from baseline if patients who are classified as responders who lose response	Female: [redacted]	[redacted]
		Male: [redacted]	[redacted]
AML	Utility value for patients who transition to AML health state	0.530 (0.053, [assumed 10% of mean])	0.426 – 0.633
Palliative care	Utility value for patients who transition to End of life health state	0.530 (0.053, [assumed 10% of mean])	0.426 – 0.633

Decrement in utilities associated with AEs (

Table 27) are calculated in the CS from the frequency of non-hematological grade ≥3 adverse events and the disutility associated with the various adverse events (CS, Table 63), taken from a range of sources, with the decrement in utilities assumed to last 4 weeks. This only used in scenario analysis. Disutilities are not considered in the base-case.

Table 27: Annual adverse event disutilities (reproduction of Table 66, CS, page 152)

Starting treatment:	Annual disutility
Fedratinib	0.001
BAT	0.003
BAT, after two JAK inhibitors	0.003
Key: BAT, best available therapy; JAK, Janus Kinase	

4.2.9 Resource use and costs

The following costs categories are included in the CS model[37] (CS,[1] Section B.3.5.).

- drug administration and acquisition costs,
- resource use associated with disease management of MF in patients treated with JAK inhibitor and BAT
- thiamine testing and supplementation
- costs associated with the management of adverse events
- costs associated with the management of AML and palliative care

4.2.9.1 Drug administration and acquisition costs

Drug acquisition costs are summarised in Table 28. Drugs are costed per opened pack. JAK-inhibitors (fedratinib, ruxolitinib) and other drugs included in BAT in the base-case are taken orally; thus, no administration costs were assumed in the base-case.

- Fedratinib

The company informed NICE that the final list price for fedratinib (and the associated patients access scheme [PAS] discount) is not yet known, but that the net price should be used and reflects the final price following discount offered to the NHS. Based on its current placeholder list price, the cost per pack of [redacted] fedratinib tablets (30 days' supply) would be [redacted].[1] The company[1] has proposed a PAS which takes the form of a simple price discount of [redacted]; taking the final net price to [redacted].

- **BAT (Comparator arm)**

BAT is comprised of a basket of treatments with the composition taken from SIMPLIFY-2[10] (n = 52) in the base-case, which is primarily composed of ruxolitinib (≈ 89%). Scenario analysis are conducted using BAT composition taken from PERSIST-2[9] (n = 98; 44.9% ruxolitinib use). No scenario analysis is conducted using the BAT composition from Schain et al.[38] (n = 37; 0.0% ruxolitinib use) or the Haematological Malignancy Research Network (HMRN)[1] (n = [redacted]; [redacted]% ruxolitinib use).

The BAT composition from SIMPLIFY-2[10] is used in the base-case justified by the company[1] (a) following their discussion with clinical experts, to reflect the composition for BAT at the point at which

patients "*should*" discontinue ruxolitinib (while patients are continued on suboptimal treatment due to the lack of targeted therapeutic options available), and (b) to be in line with the response rate and frequency of adverse events used in the base-case.

Table 28: Drug acquisition costs (adaptation of Table 70, CS, page 158)

Treatment	BAT composition (comparator arm)	BAT composition (fedratinib arm)	Pack size	Unit size	Unit type	Pack cost	Source	Duration of a pack[a]	Weekly cost[b]
fedratinib	NA	NA	[redacted]	[redacted]	Tablet	[redacted]	Net provided by the company	4.29	[redacted]
Hydroxycarbamide (hydroxyurea)	23.1%	66.7%	100	500 mg	Tablet	£9.56	eMIT	3.92	£2.44
Prednisolone	5.8%	16.7%	28	5 mg	Tablet	£0.31	eMIT	0.67	£0.47
Prednisone	5.8%	16.7%	30	5 mg	Tablet	£26.70	BNF	2.14	£12.46
ruxolitinib	88.5%	0.0%	56	5 mg	Tablet	£1,428	MIMS	4.00	£357.00
			56	10 mg	Tablet	£2,856		4.00	£714.00
			56	15 mg	Tablet	£2,856		4.00	£714.00
			56	20 mg	Tablet	£2,856		4.00	£714.00

Key: BAT, best available therapy.

[a] Excluding wastage; [b] Cost standardised per week (pack cost/ duration of a pack) - Not used in the economic model, and presented for transparency to compare weekly costs (as cost per pack in the CS model)

The company[1] further consider alternative sources to be less reliable for the following reasons (CS,[1] page 153);

- patients in PERSIST-2[9] are less comparable to JAKARTA-2[11] as JAKARTA-2 included only patients with platelet count < 100 x 10^9/L,
- Schain et al, 2019[38] is conducted in Sweden and Norway and was deemed inappropriate for a UK setting by the company's clinical experts (clarification was sought on this inconsistency. This is described in Section 4.3.4.6.1),
- the sample size of the HMRN[1] was considered too small (n=[redacted]).

Drug acquisition costs in the CS are sourced from the Monthly Index of Medical Specialities (MIMS) online database, British National Formulary (BNF) and the Drugs and pharmaceutical electronic Market Information Tool (eMIT) for drugs available in generic form (CS,[1] page 158-159). The CS[1] includes an additional 5% wastage for ruxolitinib only to account for dose adjustments, this is justified in the CS by referencing the ERG preferred base-case assumption in TA386.[13] No PAS is included for ruxolitinib.

- **BAT (fedratinib arm)**

The BAT composition for patients discontinuing fedratinib (CS,[1] page 156) was derived from SIMPLIFY-2[10], excluding ruxolitinib. BAT composition from SIMPLIFY-2[10] was re-weighted to exclude ruxolitinib.

4.2.9.2 Resource use associated with the management of MF

Resource use assumed in the CS economic model are summarised in Table 29 and comprise: A&E visits, blood tests (FBC & U&E), hospital inpatient stays, hospital outpatient visits, primary care visits, RBC unit transfusion, urgent care visits.

Resource use for each treatment in the CS is derived from: (1) the proportion of JAK inhibitors (100% fedratinib, 89% BAT comparator arm, 0% BAT fedratinib arm), (2) the baseline resource use in the absence of JAK inhibitor (assumed to be constant), and (3) relative impact of JAK use on resource use (varying with time).

Unit costs in the CS are taken from the NHS Reference Costs, Unit Costs for Health and Social Care, Private Patient Tariff and literature when appropriate.[1] Costs were uplifted where appropriate to 2019 values.

Table 29: Resource use assumed in the base-case (Table 77, CS, page 165)

Cost per week	fedratinib	BAT as comparator	BAT after fedratinib
Cost per week: 0 - 12 weeks	£210.24	£210.29	£210.66
Cost per week: 12 - 24 weeks	£190.41	£192.75	£210.66
Cost per week: 24 - 36 weeks	£170.35	£175.00	£210.66
Cost per week: 36 - 48 weeks	£170.08	£174.76	£210.66
Cost per week: 48 - 108 weeks	£169.89	£174.59	£210.66
Cost per week: 108 - 144 weeks	£150.09	£157.08	£210.66
Cost per week: 144+ weeks	£129.27	£138.66	£210.66

- **Baseline resource use**

Resource use in the CS in the absence of JAK inhibitor is taken from TA386,[13] and derived from two UK sources, ROBUST[50] and the HMRN audit (2016).[51] Scenario analysis are included using updated data from HMRN (2020).[52]

- **JAK inhibitor**

The impact of JAK inhibitor on resource use relative to BAT (excluding JAK use) in the CS is derived from HMRN (2020)[52], JUMP[13, 53] and similar assumptions to NICE TA386.[13] Scenario analyses in the CS are conducted using values from TA386[13] (derived mostly from JUMP[53]) or JAKARTA in ruxolitinib naïve patients (against placebo).[14]

4.2.10 Thiamine testing and supplementation

Additional resource use associated with thiamine testing and supplementation in the CS is included for patients receiving fedratinib only. It is assumed in the CS that thiamine testing occurs at baseline, then once every month for the first 3 months, then once every 3 months, and assumed to be conducted alongside other routine tests. A cost of £31 per test[54] is assumed to reflect that tests have to be sent to external laboratories for processing as few centres in the UK have the capacity to process thiamine tests.

In the CS, [redacted]% of patients are assumed to require thiamine supplementation based on JAKARTA-2[11] (n = [redacted]). Patients are assumed to incur the cost of a full 90-day course (assumed to be 200 mg daily) upon initiation of fedratinib and discontinuation.

4.2.11 Costs associated with the management of adverse events

Adverse event unit costs (CS,[1] page 167) in the CS are taken from a multitude of sources including the National Health Service (NHS) Reference Costs, Unit Costs of Health and Social Care, assumptions used in TA386 and literature when necessary.[1]

The annual frequency of grade ≥ 3 adverse events in the CS is then multiplied by the respective unit costs to obtain an annual cost for managing AEs (Table 30) in patients treated with fedratinib, BAT (including ruxolitinib) and BAT (excluding ruxolitinib) following fedratinib discontinuation.

Table 30: Annual AE costs (reproduction of Table 80, CS, page 168)

Treatment	Annual cost
fedratinib	£27.09
BAT as comparator	£98.83
BAT after fedratinib	£88.79

4.2.12 Other costs

The weekly cost associated with the management of AML in the CS is derived from Wang et al (2014),[55] before being uplifted to 2019 values and is assumed in the CS to be £32,087 per year. The weekly cost associated with the management in palliative care (the last 8 weeks prior to death) in the CS is derived from Round et al (2015),[56] again before being uplifted to 2019 values and was assumed to be £813.72 per week.

4.2.13 Model validation

The CS[1] (Section B.3.10.1) describes a number of measures taken by the company to validate key assumptions in the economic model during an advisory board[27, 37] and also to verify the executable model.

The CS[1] states that clinical and health economics experts at the advisory board were asked to validate the following elements: model structure, clinical care pathway, relevant comparator, extrapolation for OS and the composition for BAT.

The CS[1] also describes technical validation by a programmer who was not involved in building the model in terms of reviewing the programming code and assessment of the behaviour of the model results to changes in inputs.

4.2.14 Results

4.2.14.1Deterministic base-case

Table 31 presents the mean estimates of cost-effectiveness generated using the company's model for the comparison of fedratinib versus BAT in the subset of patients with intermediate-2/high risk MF (1,000 patients).

The deterministic version of the company's model suggests that fedratinib is expected to generate an additional [redacted] QALYs at an additional cost of [redacted] per patient compared with BAT. The corresponding ICER is £13,905 per QALY gained. The probabilistic version of the company's model suggests that fedratinib is expected to generate an additional [redacted] QALYs at an additional cost of [redacted] per patient compared with BAT; the corresponding ICER is £10,384 per QALY gained.

Table 31: Base case results (based on net price) – reproduction of table 176, CS

	Total costs (£)	**Total LYG**	**Total QALYs**	**Inc costs (£)**	**Inc LYG**	**Inc QALYs**	**ICER vs baseline (£/QALY)**
Deterministic results							
BAT	[redacted]	[redacted]	[redacted]	-	-	-	-
fedratinib	[redacted]	[redacted]	[redacted]	[redacted]	0.848	[redacted]	13,905
Probabilistic results							
	[redacted]	[redacted]	[redacted]	[redacted]	-	[redacted]	-
	[redacted]	[redacted]	[redacted]	[redacted]	0.898	[redacted]	10,384

Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.

4.2.14.2Company's PSA results

Figure 15 presents the cost-effectiveness plane, and Figure 16 presents the CEACs for fedratinib versus BAT. Assuming willingness-to-pay (WTP) thresholds of £30,000 and £50,000 per QALY gained, the company's results suggests that the probability that fedratinib generates more net benefit than BAT is [redacted] and [redacted] respectively.

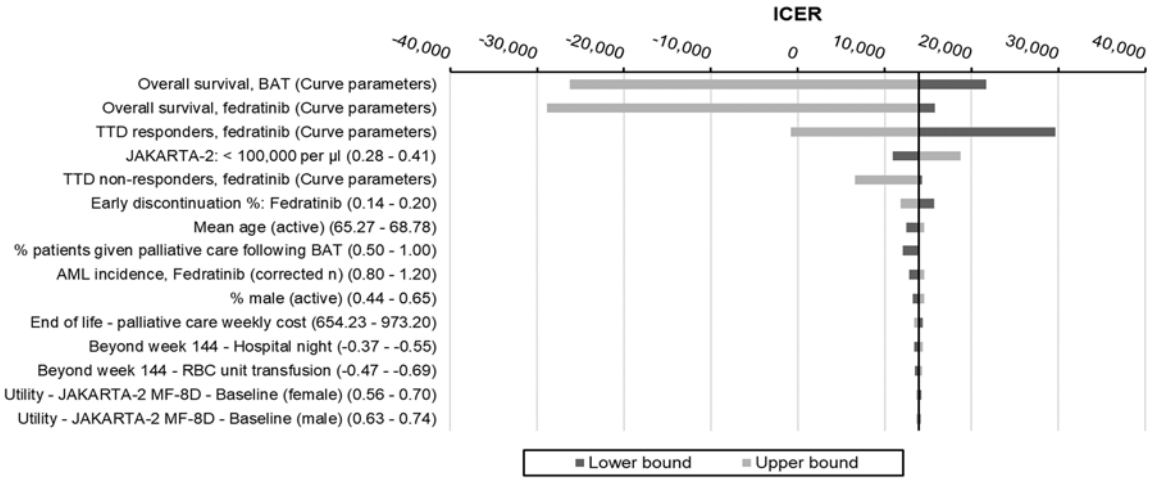
Figure 15: Cost-effectiveness plane – fedratinib vs BAT (Reproduction of Figure 28, CS, page 177)

Figure 16: Cost-effectiveness acceptability curve – fedratinib vs BAT (Reproduction of Figure 29, CS, page 177)

4.2.14.3Company's One-way Sensitivity analysis

The company's tornado plot is shown in Figure 17. The plot indicates that OS, TTD for fedratinib and the distribution of platelet count are the key drivers of the ICER for fedratinib versus BAT.

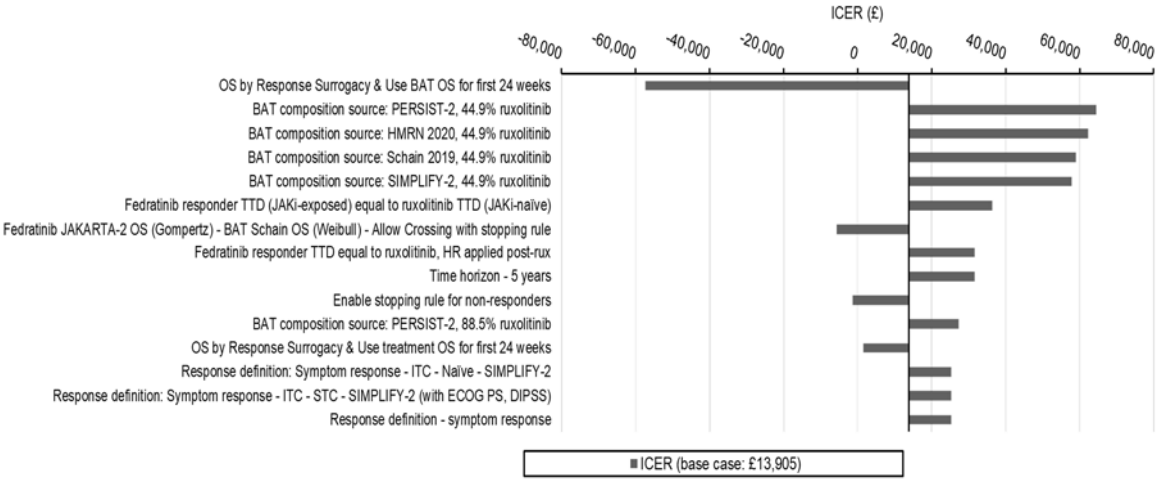
Figure 17: Results of one-way sensitivity analysis (Reproduction of Figure 30, CS, page 178)



4.2.14.4Company's scenario analysis results

The results of the company's scenario analyses are summarised in Figure 18. ICERs were very sensitive to the different scenarios examined by the company and varied from fedratinib being dominant (surrogacy scenario) to generating an ICER of £64,445 per QALY gained (when using the BAT composition from PERSIST-2[9]). Of note, the proportion of ruxolitinib use from PERSIST-2[9] (44.9%) is used for all scenarios examining different BAT composition (other treatment part of BAT).

Figure 18: Summary of modelling scenarios which had the most impact on the base case ICER (Reproduction of Figure 31, CS, page 180)



4.3 Critique of company's submitted economic evaluation by the ERG

The critical appraisal presented in this section focuses on the key issues identified by the ERG; less important issues and less substantial sources of potential bias are not fully critiqued here. The ERG assessment is also limited by the information made available to the ERG by the company during both the submission[1, 37] and clarification stage.[18, 29]

The ERG has identified a number of inconsistent statements between the CS[1] and advisory board notes[27] sent as part of the CS reference pack. This issue was raised with the company by the ERG at the clarification stage.[18] In response, the company submitted an updated version of the advisory board notes[37] considered to be the final version. While both documents are broadly similar, with the final version including a few additional sentences and graphs (extrapolation for OS), the ERG notes some discrepancies between the two versions. As the ERG is not able to confirm the reasons for these discrepancies (except for the fact that one version is final, and the other one was a draft version), for transparency and completeness, the ERG highlights in this section the discrepancies it considers relevant, between the final[37] and draft version[27] of the advisory board notes.

In summary, while the ERG recognises the challenges arising from the single arm nature of JAKARTA-2,[11] the impact of the clinical hold and the paucity of evidence for the comparator assumed in the economic model, the ERG has a number of concerns regarding the company's implemented economic model.[37] While some of these issues are debatable and may reflect matters of subjective opinion, others reflect more serious underlying problems regarding the conceptualisation and implementation of the model and the use of evidence to inform the model's parameters. Consequently, the ERG has serious doubts regarding the validity of the results presented within the CS[1, 37] and would advise considerable caution in their interpretation and use in informing decision-making. These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethinking of the model's logic. As such, the impact of these issues on the expected cost-effectiveness of fedratinib is not clear.

While the ERG undertook some exploratory analysis to explore the impact of some of these issues on the ICER for transparency and completeness, the ERG advises caution, as these analyses were limited by the company's model structure, but also evidence available.

4.3.1 Methods for reviewing the company's economic evaluation and health economic model

The approach taken by the ERG to appraise the company's economic analysis[1, 37] consisted of the following key steps:

1. scrutiny and understanding of the logic of the VB code.[37] While the VB code was checked for potential errors, it was not possible to carry a full line-by-line review of the VB code for each component of the model due to time and resource constraints and more broader conceptual issues described later,
2. double-programming of some functions in Excel directly. This allowed to check that there were any serious programming errors within the VB code,
3. comparison of the model's predictions[37] against inputs, to ensure no serious programming errors are present,
4. ensuring that changes in inputs changes results appropriately (predictive validity),
5. validation of key assumptions using clinical experts,
6. assessment of differences with TA386,[13] as the company makes a number of references to this appraisal to justify their model approach,
7. check of inputs and/or assumptions[37] against the references cited by the company (literature for key inputs and advisory board notes for key assumptions).[1]

Overall, the company's Excel economic model[37] is relatively transparent, with clear notation and references for inputs provided. Key calculations (health state sojourn time, assignment of costs, QALYs) in the company's model[37] are undertaken directly within VB for Microsoft Excel, and the VB code is clearly described and well annotated. The company's model includes a number of options, most of them are relevant (although some lack face-validity and are not implemented correctly as described in Section 4.3.3). A minor comment by the ERG, is that options for the selection of different parametric extrapolations could have been more transparent and clearly set out in the "control" sheet directly, rather than forcing the user to access individual worksheets for each specific survival outcome (OS, TTD, DoR). The economic model also did not include all data needed for validation. For instance, some KMs were missing, or those available for TTD or DoR did not allow validation of predictions (as the company subsequently stated at the clarification stage[18] that those present in the model were censored for death). These were requested and supplied at the clarification stage.[18] It is the ERG's view that should all information be included in economic model for validation, errors identified in Section 4.3.4.2.2 by the ERG would have been avoided.

4.3.2 Adherence of the company's model to the NICE reference case

The company's economic analysis[1, 37] is generally in line with the NICE Reference Case (see Table 32).[57]

Table 32: Adherence of the company's economic analyses to the NICE Reference Case

Element	Reference case	ERG comments
Defining the decision problem	The scope developed by NICE[2]	The company[1] focuses its submission on adults with MF previously treated with ruxolitinib. No analysis is conducted in (a) people with no prior treatment with ruxolitinib or (b) for whom ruxolitinib is not appropriate. It is not entirely clear from the CS[1] the population entering the model. The ERG interpretation (from the BAT composition assumed) is that the population focuses mostly on patients who are relapsed, refractory or intolerant to ruxolitinib who are continued on ruxolitinib treatment (e.g., did not discontinue) who would switch to fedratinib (assumed to be 89% of the population) but also include a minority of patients who discontinue ruxolitinib and receive a different BAT treatment.
Comparator(s)	As listed in the scope developed by NICE[2]	The comparator in the final NICE scope[2] (for patients previously treated with ruxolitinib) is established clinical practice (including but not limited to hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin and red blood cell transfusion). The company's model consider best available therapy (BAT) as the comparator, including high ruxolitinib use (89%). This is reasonable if the majority of the population entering the economic consist mostly of people who are continued on ruxolitinib who would switch to fedratinib and not people who discontinue ruxolitinib. Both the population and comparators needs to be consistent.
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Health gains accrued by patients are valued in terms of QALYs gained. Impacts on caregivers are not included.

Element	Reference case	ERG comments
Perspective on costs	NHS and PSS	Although not directly stated, the analysis adopts an NHS and PSS perspective.
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The results of the company's base case analysis are presented in terms of the incremental cost per QALY gained for fedratinib versus BAT.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The model[37] adopts a 30-year time horizon in this base-case.[1] At this timepoint, virtually all patients in the model have died.
Synthesis of evidence on health effects	Based on systematic review	Relative treatment effects for response rates were estimated using MAIC/STC based on the company's systematic reviews; concerns are described in Section 4.3.4.7. Overall survival (OS) for patients initiating fedratinib is taken from the JAKARTA-2 trial.[11] Whilst OS in patients initiating BAT is identified from a systematic review of literature conducted in patient who discontinued ruxolitinib; concerns are described in Section 4.3.4.6. OS from JAKARTA-2 is compared naively against Schain et al[38]
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	Health state utility values are based on MF-8D,[46] which is a condition-specific preference based measure in MF developed and used in TA386.[13] The MF-8D is derived from the MFSAF and EORTC QLQ-C30. The use of the MF-8D is justified by the company[1] owing to the inability of the EQ-5D to capture key symptoms in MF.
Source of data for measurement of HRQoL	Reported directly by patients and/or carers	
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	

Element	Reference case	ERG comments
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	No additional equity weighting is applied to estimated QALY gains.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Resource costs include those relevant to the NHS and PSS. Unit costs were valued at 2019 prices.
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	Costs and health effects are discounted at a rate of 3.5% per annum.

BAT – Best Available Therapy; NHS - National Health Service; PSS - Personal Social Services; QALY - quality-adjusted life year; HRQoL - health-related quality of life; EQ-5D - Euroqol 5-Dimensions; EORTC QLQ-C30 - European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-C30; MF – Myelofibrosis; OS – Overall Survival

While the population entering the economic model[1, 37] aligns with the NICE final scope (a subset of it),[2] it is not entirely clear from the CS,[1] what the population is that are entering the model, at what time point patients enter the model, and whether the submission focuses only on: (a) patients who are relapsed, refractory or intolerant to ruxolitinib who are continued on ruxolitinib treatment and who would switch to fedratinib or (b) includes a mix of patients who discontinue ruxolitinib and would start fedratinib.

This is an important distinction as the comparator (BAT composition; proportion of ruxolitinib use) will vary according to the population considered (key driver in the economic model). The company was asked to clarify the population entering the economic model. In response to clarification[18] (see clarification question B3[18]) the company stated that "*the population entering the model are patients with intermediate-2 or high-risk primary or secondary MF, who have been treated with ruxolitinib and are refractory/relapsed or are intolerant to ruxolitinib*".

The company[1] assumes that 89% of patients in the comparator arm receive ruxolitinib based on the BAT composition in SIMPLIFY-2.[10] Consequently, the ERG's interpretation of the CS[1] is that the population entering the economic model[37] consist mostly of people that are continued on ruxolitinib while achieving a suboptimal response (at the point where they become resistant/intolerant rather than when they discontinue ruxolitinib). In particular, the CS (CS,[1] page 155) states that "*If fedratinib was available, there would be an opportunity to switch patients who are relapsed, refractory or intolerant to ruxolitinib onto an effective therapy. In current practice, if the composition for BAT were defined from the point at which patients should discontinue ruxolitinib, the UK clinical experts indicated they would expect ruxolitinib use in this population to be similar or greater than that used in SIMPLIFY-2 (89%)*."

Clinical advisors to the ERG reported that patients are often kept on ruxolitinib as long as possible despite not achieving criteria for full response and achieving suboptimal response. This is done because patients would typically derive some benefits (notably for symptom control and survival) and as there are no other effective alternative therapy available after ruxolitinib.

The comparator defined in the NICE final scope[2] is established clinical practice (including but not limited to hydroxycarbamide, other chemotherapies, androgens, splenectomy, radiation therapy, erythropoietin and red blood cell transfusion). The ERG considers the inclusion of ruxolitinib to be relevant if the population entering the economic model is patients who continue ruxolitinib while achieving a suboptimal response and who would switch to fedratinib. For patients who discontinue ruxolitinib (because of an AEs or lack of efficacy), ruxolitinib is not a relevant comparator. This is difficult to disentangle from the CS as the population is pooled together. As described in Section 4.3.4.1,

given the evidence available, and the discrepancies in the company's economic model, the ERG believes that the population needs to be separated onto patients who are continued on ruxolitinib (ruxolitinib being the comparator) and patients who discontinue ruxolitinib (BAT without ruxolitinib being the most appropriate comparator). The ERG believes that such approach would help with interpretation of results. The comparators are also different according to these two populations.

The ERG is also unclear whether the population eligible for fedratinib in the UK would be closer to patients initially recruited in JAKARTA-2,[11] or those included as part of the re-analysis of JAKARTA-2.[23] This is unclear from the CS.[1] The economic model[37] uses data from the original analysis;[11] not the re-analysis using a more stringent definition of resistant/intolerance.[23, 24]

Compared with the NICE reference case, health state utility values are based on the MF-8D,[46] rather than the EQ-5D. This is justified due to psychometric properties of the EQ-5D in MF.[44, 47] The ERG notes that this is consistent with TA386[13] for the assessment of ruxolitinib in myelofibrosis. However, while the ERG recognises concerns with the psychometrics properties of the EQ-5D for this population, the ERG believes that a scenario analysis using the EQ-5D should be presented for transparency and completeness in line with the NICE reference case.[57]

4.3.3 ERG Critique of the modelling performed by the company

4.3.3.1 Model verification

Given the time taken to assess the broader conceptual concerns described in Section 4.3.4, the ERG was not able to fully verify the model. However, during the model verification process (in particular checking predictions from the model), the ERG identified a number of programming/conceptual errors in the implementation of both the AML and palliative care health state; these are discussed in further details in Section 4.3.4.2.5. Whilst these issues are minor in relation to other conceptual issues identified, the ERG believes that they reflect a poor modelling choice.

4.3.3.2 Correspondence of the model inputs and the original sources of parameter values

Where possible, the ERG checked the company's model inputs against their original sources, although many of these were drawn from unpublished analyses of JAKARTA-2.[11]

Given the time taken to assess the broader conceptual concerns described in Section 4.3.4, the ERG focused on key inputs and therefore, cannot confidently confirm that all the company's model inputs matches their original sources. Nevertheless, no discrepancies were identified by the ERG for the inputs that were checked.

4.3.4 The main issues identified by the critical appraisal

The main issues identified from the ERG's critical appraisal are summarised in Box 1. These are discussed in further detail in the subsequent sections.

Box 1: Summary of the main issues identified within the company's health economic model

1. Concern regarding alignment between the population, evidence used and cost (BAT composition),
2. Concerns regarding the modelling approach,
3. Concerns regarding the lack of consistency between assumptions made for the comparator and the intervention arm,
4. Concerns regarding the selection process for OS survival model in patients initiated on fedratinib,
5. Concerns regarding evidence used for OS for the comparator and resulting predicted survival,
6. Concerns with using response rate and its application in the economic model,
7. Concerns with using duration of response and application in the economic model,
8. Lack of face-validity for the scenario analysis assuming surrogacy between spleen and survival,
9. Lack of face validity for the scenario using the stopping rule,
10. Concerns regarding the company's cost assumptions,
11. Concerns regarding the survival model selection process for TTD,
12. Concerns regarding the company's HRQoL assumptions,
13. Inclusion of adverse event in the model,
14. Uncertainty regarding resource use.

4.3.4.1 Concern regarding alignment between the population, evidence used and cost (BAT composition)

Key drivers in the economic model[37] are: (1) the BAT composition (e.g proportion of ruxolitinib use), (2) overall survival and (3) how long patients are kept on treatment.

The analysis presented in the CS[1] is for a combined population. The ERG notes that because of the single-arm nature of the trial,[11] and evidence used for the comparator arm in the economic model (derived from SIMPLIFY-2[10] and Schain et al.[38]), there are a number of inconsistencies and the ERG believes that for both clinical and economic reasons, separating the populations could help with both the interpretation of results and the underlying assumptions. The ERG believes that, should evidence be taken from a RCT for both the intervention and comparator (e.g., conducted in the same population),

that separating the population becomes less necessary, although still relevant as key outcomes (such as overall survival) will be affected by the proportion of ruxolitinib use and also the population considered (resistant versus intolerant). Comparators are also different according to the population considered, making it relevant to separate each population.

According to the ERG's interpretation of the CS,[1] the company consider a mix of patients that are continued on ruxolitinib (89%) at the point where ruxolitinib "should" be discontinued and patients that no longer receive ruxolitinib (11%).

The ERG is unclear about the population targeted in the CS[1] and economic model.[37] Consequently, the company was asked (see clarification response,[18] question B3) to clarify (a) what is the population entering the model, (b) to clarify how inputs in the model match the population considered and (c) to clarify the following statement in the CS (CS,[1] page 98): "*ruxolitinib alone is not considered a relevant comparator but is instead included as part of the basket of treatments in BAT. This attempts to ensure alignment between costs and efficacy inputs (See Section B.3.6.1)*".

In its response, the company provided a table describing the source for key inputs in the economic model[37] and stated "*as shown in the table above, the proportion of patients who respond in the BAT is informed by an ITC between JAKARTA-2 and SIMPLIFY-2 trials. The adverse event data and the composition of BAT was also taken from SIMPLIFY-2. Utilities are impacted by the proportion of patients who respond and the frequency of adverse events. Costs are impacted by BAT composition and adverse events. By using the same source across these inputs, we attempt to align the costs and efficacy inputs for the BAT arm*".

The ERG remains unclear and finds the response from the company to be unsatisfactory for the following reasons:

- OS is a key driver in the economic model[37] and is not aligned with costs (BAT composition). In the BAT arm in the economic model, 89% of patients are assumed to be on ruxolitinib continuation, but OS is taken from a study post-ruxolitinib cessation (patients are no longer on ruxolitinib),
- response rates used in the economic model are unreliable because of significant differences between the population recruited in JAKARTA-2[11] and SIMPLIFY-2[10] which have not been adjusted for in the unanchored ITC conducted by the company. SIMPLIFY-2 recruited patients who had myelofibrosis and previous ruxolitinib treatment for at least 28 days who either required red blood cell transfusions while on ruxolitinib or ruxolitinib dose reduction to less than 20 mg twice a day with at least one of grade 3 thrombocytopenia, anaemia, or bleeding at

grade 3 or worse, with palpable spleen of at least 5 cm and without grade 2 or greater peripheral neuropathy. Inclusion criteria were different in JAKARTA-2. Not all of these differences in inclusion criteria were accounted for in the unanchored ITC conducted by the company. There are also issues described in Section 4.3.4.7

- response rates for BAT (in people receiving ruxolitinib or not) can be derived from the key SIMPLIFY-2[10] publication (Figure 2 and Figure 2). Therefore, it is possible to estimate response rates for people on ruxolitinib and BAT without ruxolitinib separately,
- while response rate for BAT in patients initiated in the comparator arm aligns with the BAT composition; the response rate for BAT in patients initiated on fedratinib does not align with the BAT composition (as the same response rate [comprising 89% ruxolitinib] is used, but assuming no costs for ruxolitinib). Therefore the ERG consider that costs are only partially aligned with response rates in the economic model.[37] Patients initiated on fedratinib who move to BAT, also receive BAT after 2 JAK inhibitors (which also does not align with the SIMPLIFY-2 population),
- finally, the ERG believes that because the company assumes that the large majority of patients (89%) in BAT receive ruxolitinib, it is appropriate to examine ICERs against ruxolitinib alone as a direct comparator. The ERG further notes that the CS (CS,[1] page 155) reports the proportion of ruxolitinib in the UK population to be higher than that reported in SIMPLIFY-2. Therefore, the ERG believes that ruxolitinib alone should be considered a comparator separately.

Consequently, the ERG believes that given the discrepancies in the company's economic model[1, 37] between key efficacy inputs (OS) and costs (BAT composition), and the fact that comparators are different according to the population of interest, two analyses should be presented separately for the following comparators and population;

(1) [redacted]Ruxolitinib continuation in people previously treated with ruxolitinib who remain on treatment while achieving a suboptimal response who could switch to fedratinib (when "should" have discontinued),
    a. This appears to be the key population targeted in the CS (as it is assumed that 89% receive ruxolitinib). By definition, ruxolitinib alone (or combination) is a *de facto* comparator for this population.
(2) BAT (excluding ruxolitinib) in people previously treated with ruxolitinib who discontinue ruxolitinib treatment (due to AEs or efficacy reason – representing 11% of the population in the CS) and who would move to fedratinib.
    a. While BAT without ruxolitinib is perhaps the most relevant comparator for this population, patients could "in theory" receive ruxolitinib again (ruxolitinib rechallenge), although this is rare in UK practice.

The ERG believes that separating the population would allow a better alignment between costs and other inputs in the economic model but also facilitate interpretation of results (are patients taken at the point when they "should" discontinue ruxolitinib or at the point of ruxolitinib cessation). At present it is challenging to disentangle assumptions when a mix population is used. It is the ERG's view that separating the population would not require more assumptions compared with those already made in the CS. The ERG further notes, when conducting scenario analysis for different BAT proportions, no scenario analysis is conducted by the company using the BAT composition from Schain et al.[38] (n = 37; 0.0% ruxolitinib use).

Finally, the ERG notes that the BAT composition (proportion of ruxolitinib use) and OS is also likely to be different according to the population considered (refractory, relapsed or intolerant). The ERG requested the company to provide the KM graph for OS for the three subgroups (see clarification response,[18] question B10). While the p-values were non-significant, the ERG notes clear visual differences, indicating possible differences between these subgroups.

4.3.4.2 Concerns regarding the modelling approach

The ERG does not believe the modelling approach taken by the company to be appropriate or robust. The ERG questions: (1) the value for using an individual-based approach and (2) the value for separating patients into group categories.

While the ERG understands that an individual-event based approach (albeit a very different implementation) was used in TA386,[13] the level of evidence for the effectiveness of the intervention is markedly different (two RCTs informed efficacy data in TA386[13] vs. a single arm trial[11] in this appraisal). A more detailed comparison is provided in this section 4.3.4.2.1.

The ERG is concerned that the CS[1] trades-off robustness and transparency for its base-case in favour of a perhaps more flexible model (to mimic the approach in TA386; albeit the models are very different in their implementation), which also fails to address these structural uncertainties, when evidence is already limited, and thus not capturing the flexibility it was designed for.

The ERG agrees with the company[1] that an individual-based approach, in general, provides additional flexibility over a cohort model and recognises the value for flexibility when assessing important relevant scenarios and structural uncertainties. However, the current implementation in the CS[37] introduces a number of biases, and mathematical and conceptual errors, with results for relevant scenarios (such as examination of the impact of the stopping rule, estimation of OS through surrogacy relationship and including the worsening in quality of life) lacking face-validity as described in Section 4.3.4.3.1, Section 4.3.4.9 and Section 4.3.4.10 respectively.

The ERG also questions the value for separating patients into group categories, when evidence is already limited and a stopping rule is not in place for fedratinib (although the ERG acknowledges that this is debatable and it is unclear whether this would apply in practice). Consequently, a number of structural assumptions are made introducing further uncertainties and biases as described in Section 4.3.4.2.

4.3.4.2.1 Comparison with TA386

The company makes a number of references to TA386[13] to justify its modelling approach. Consequently a comparison of the two models is presented for transparency. The ERG notes that whilst both models are individual-event based, their implementation is markedly different. As described in TA386,[13] an individual-patient based model was used to model the progressive nature of MF (worsening in HRQoL in the supportive care health state) and explore the impact of different structural assumptions. As described in TA386,[13] a stopping rule is also included in the base-case as this was part of the licensing for ruxolitinib.[58] However, more relevant, described in TA386,[13] overall survival is estimated as a function of response and the time in previous health state, and therefore using an individual based-approach also allowed the model to deal with time-varying transition probabilities (in addition to modelling the progressive nature of MF and exploring other structural uncertainties).

In contrast, the CS base-case[1, 37] does not consider either the progressive nature of MF (worsening in HRQoL), nor a stopping rule for fedratinib; although these are explored in scenario analysis (but lack face validity - a description of their limitations is presented in Section 4.3.4.3.1 and 4.3.4.10 respectively).

Perhaps more important and relevant, is that OS in the CS[1, 37] is calculated independently from response and the time spent in previous health states (estimated from the time patients enter the model). Therefore, compared with TA386[13] where changes in the response rate affected survival (similar concept as with the state-transition model), in the company' model,[1, 37] OS is fixed irrespective of other inputs (similar concept as with the partitioned survival model). The company's model therefore does not consider subsequent time varying probabilities (OS is sampled from model entry, rather than estimated as function of response or time in previous health state).

The ERG therefore questions the value for an individual based-approach when OS is modelled independently from response and the time in previous health state, and it does not use individual characteristics to define patient pathways. The modelling approach in the CS (independent modelling for OS) led to biases for the estimation of the time on treatment which are described in Section 4.3.4.2.2.

It should be noted that because the model is unnecessarily complex (when evidence is already lacking), the scope for error is increased as shown in this section.

In the CS, patients are also separated into responders and non-responders, and this is used in the company's model[1, 37] to estimate the time on treatment according to the response categories and to assign utility values. First, the process to separate patients into the group categories is different to TA386.[13] As described in TA386,[13] patients are primarily split into patient groups to accommodate the stopping rule as part of ruxolitinib licensing.[58] In the company's economic model,[1, 37] response does not affect overall survival and therefore, while the ERG is of the view that this scenario lacks face validity, there is no value in separating patients to examine the effect of the stopping rule as implemented by the company (examined in scenario analysis). Splitting patients in the CS also resulted in additional assumptions for TTD (described in Section 4.3.4.2.2), which has the effect of increasing the uncertainty when evidence is already limited. While the ERG considers that separating patients, is in principle appropriate, the ERG believes that there is a trade-off between the value for separating patients, and additional assumptions and uncertainties introduced that are required to accommodate this. In particular the amount of evidence is markedly different between TA386[13] (based on 2 RCTs) and this appraisal (one single arm study[11]). The company's economic model is also not a true response-based model, as response does not affect OS. There are also a number of concerns about the reliability of estimation of response which are described in Section 4.3.4.7.

4.3.4.2.2 Biases in the estimation of TTD by sampling OS and TTD independently from each other

As previously described, in the CS economic model[37] OS, TTD and DoR are sampled independently from each other, and therefore TTD (and DoR) is truncated by OS. This leads to inconsistencies for TTD between the model predictions and the TTD survival function used for that particular event reported in the CS[1] as shown in

Figure 19 (with transition to AML set to “No” [Sheet “Control”, Cell J218] and assumption of 0% receiving palliative care) generated by the ERG (assuming 10,000 patients).

Figure 19: Comparison of prediction for TTD for responders and non-responders against the distribution used for that particular event (generated by the ERG)

The ERG sought clarification[18] from the company on this issue (see clarification response,[18] question B4). In response, the company stated that death was treated as censored when estimating TTD and DoR and therefore, the discrepancy between the distribution reported in the CS[1] and final predictions in the economic model[37] was intended (to avoid double counting). The ERG notes that there is no mention in the CS[1] that death was censored in TTD or DoR.

While the ERG agrees in principle with the general approach taken by the company; e.g., censoring death in TTD (to avoid double counting), because of the modelling approach (independent modelling but also separation of patients), there is a discrepancy between the hazard of death used in the economic model (which acts as a competing risk) and the hazard for treatment discontinuation estimated from the data (estimated in the subset of responders or non-responders). In simple terms, TTD in the economic model[37] is estimated in patients alive at 24 weeks, with a response or not. In contrast, the hazard of death in the economic model is taken from the overall population including early death, early discontinuer, responders and non-responders. There is therefore a mismatch.

The ERG further notes that when selecting TTD (Section 4.3.4.12), the company selected distributions citing clinical plausibility and selected distributions that provide the shortest time on treatment.[1] However, as death is supposed to be censored in TTD and assumed to be acting as a competing risk in the CS model,[37] it is the ERG's view that the model selection for TTD becomes less logical (as

distributions can no longer be excluded if there is a plateau – as they would be truncated by death using the company approach – further details available in Section 4.3.4.12).

Consequently, in light of new information provided by the company in the original clarification questions,[18] and issue described above, the ERG asked further clarification questions[29] to the company (see clarification response,[29] question A1 & A2) and requested the company to provide data on how many patients were censored for death in TTD, the KM for TTD (including death as events) for (1) the overall population (not split by group), (2) responders at 24 weeks and (3) non-responders at 24 weeks. This was requested by the ERG in order to validate model's predictions for TTD, as none were presented in the CS or economic model. In its response,[29] the company stated that "w*hile the code for TTD was set up to censor for death after 24 weeks, upon investigation, it was found that no deaths were recorded which led to censoring for TTD after 24 weeks. Therefore, the original concern of the ERG holds true that the TTD in the model predictions will be lower than the TTD parametric function*".

The response from the company confirm the original concern from the ERG that the modelling approach taken (individual based-approach as implemented in the CS) to be inappropriate and generate biases (

Figure 19). Given its modelling approach (independent modelling of OS and TTD),[37] this issue cannot be resolved robustly.

4.3.4.2.3 Additional uncertainties introduced as a result of separating patients into group categories

Clarification[18] was sought from the company by the ERG on the value of separating responders and non-responders in the model if no stopping rule is assumed and OS is modelled independently of response status (see clarification response,[18] question B2).

In its response, the company stated that this allowed: (1) a utility benefit for responders, (2) the application of TTD is based on response and (c) this allowed amending the model as per ERG request for clarification question B6.[18] The ERG agrees that splitting patients into responders and non-responders allows assignation of different utility values according to the level of response, but the ERG still has concerns which are described below. The ERG notes that regarding point (2), the approach for estimating TTD was influenced by the approach to modelling (separating patients) rather than the other way around and therefore does not consider this argument to be valid. Regarding point (3), the ERG notes that model choices were made prior to the ERG assessment and therefore this is not a valid argument.

The ERG does not believe that the benefit in separating patients in the model only to account for differences in quality of life to outweigh the adverse consequences and additional uncertainties introduced, when the CS evidence is already immature and limited, and alternative methods (although perhaps more crude) could have been employed to capture differences in quality of life between treatments. The ERG's key concerns are summarised below:

1 The ERG has a number of concerns about the reliability of the response rate (described in Section 4.3.4.7) used in the company model[1, 37] and therefore does not believe that response rate should be used in the model, or at least should not drive the model choices.

2 A key driver in the model is TTD. However, by separating TTD onto responders and non-responders, the sample size and number of events for each group become smaller (n=[redacted] for responders and n=[redacted] for non-responders), increasing the uncertainty. This can be observed by the large variation in parametric extrapolation. This also led the company to arbitrarily use TTD in patients with no-spleen response as a proxy for the TTD for non-responders according to spleen and symptoms. The CS[1] states: "*As the number of patients who were receiving fedratinib at 24 weeks but did not have spleen or symptom response (and did not have a censored TTD) was only* [redacted]*, this was insufficient to estimate parametric curves for TTD. Therefore, the spleen non-responder TTD was used instead (number at risk =* [redacted]*)*".

3 Because TTD is estimated as a function of response, TTD predicted in the economic model[37] is different according to the response definition. To illustrate this, the ERG generated the TTD

survival functions for pooled responders and non-responders at 24 weeks (using survival functions selected by the company). The generated TTD (removing AML transformation and transition to palliative care – 10,000 patients sampled) is presented in Figure 20. In summary, TTD (for responders and non-responders at 24 weeks) predicted using the company's response definition (spleen or symptom – red line) is different compared with if TTD was predicted using symptom response only (green line). Because TTD is a driver in the model, this would affect the ICER.

Figure 20: TTD estimation using different response definition

4 Finally, whilst the approach taken by the company allows an explicit separation of responders and non-responders, because of the modelling approach (OS modelled independently from response and TTD), the predicted time in health states is not accurate as described in Section 4.3.4.2.4. Furthermore, as described in Section 4.3.4.13, there are also concerns in the way utility values are applied in the economic model. In particular the assumption that utility values for responders and non-responders for non-JAK treatments (BAT after fedratinib [100%]; and the small proportion of BAT for the comparator [11%]) is the same as the utility values for JAK treatments (fedratinib or ruxolitinib).

4.3.4.2.4 Conceptual inconsistencies that responders have the same survival as non-responders

The ERG notes that the general approach taken by the company to separating patients led to conceptual inconsistencies with the CS argument.[1] Given the disconnect between OS and response rate, the time to death is estimated in the CS to be the same for responders and non-responders.

This is in contradiction with the CS[1] argument (CS,[1] page 48) which states that "*Feedback received from clinicians indicates there is clinical plausibility for the use of SVR as a surrogate marker for survival. The SVR outcomes observed in JAKARTA and JAKARTA-2, taken together with the OS benefit versus placebo observed in JAKARTA, further support the idea that fedratinib offers a survival benefit to patients treated with ruxolitinib*".

Whilst this could be plausible (should it be considered that the model generate conceptually valid predictions), this therefore does not support the assumption of differential survival between treatments.

Additional clarification was sought from the company (see clarification response,[29] question A1) to provide the KM for OS for responders (n=[redacted]) and non-responders (n=[redacted]) from 24 weeks. The ERG notes that the KM for non-responders provided by the company is not correct, as the KM includes [redacted] patients (n=[redacted] events, n=[redacted] censored) despite this subgroup based on only [redacted] patients (or [redacted] if using spleen definition). Therefore, the ERG only presents the KM for responders (n=[redacted]) in Figure 21 and compares this against the model predictions for both responders and non-responders. While the ERG recognises the small sample size as well as the large number of censored patients, it is clear from Figure 21 that in the company's economic model[37] the predicted times to death after week 24 in patients initiated with fedratinib that are responders or non-responders are the same, but also the predicted OS for responders does not align with the KM survival functions. This therefore questions the robustness of the predicted time in each health state and the predicted quality of life.

Figure 21: Comparison of prediction for OS from week 24 for responders against the KM

[redacted]

4.3.4.2.5 Programming and conceptual errors

Following review of the CS model predictions, the ERG identified a number of programming and conceptual errors, some identified at the clarification stage. The ERG was not able to correct all of those within the time available.

In the company base-case,[1, 37] survival for patients with AML is re-estimated using a formula described in Appendix L.8 of the CS.[37] The ERG notes that when running the company's base case model[37] for 10,000 patients, the model predicts that [redacted] patients in the fedratinib arm would transform to AML, of whom [redacted] would have a predicted time in AML of less than zero, which is not possible. Patients who are predicted to experience AML, with a time less than 8 weeks are also assumed to move directly to palliative care and therefore have a time in health state equal to zero. The time in the AML health state is estimated to be zero for [redacted] patients ([redacted]%) with AML (while the ERG understands this was intentional, this highlight some conceptual issues to the ERG).

The median (mean) predicted time in the AML health state for the [redacted] patients with a time greater than 0 was [redacted] (mean: [redacted]) weeks (which seem reasonable when looking at Mesa et al[42]). When the options for re-estimating survival is disabled, the model predicts a median (mean) time in the AML health state that is considerably higher (median: [redacted] weeks; mean: [redacted] weeks), which lacks face validity.

A programming error was also identified in that the model predicts that some patients enter the palliative care health state despite setting the proportion to zero. Indeed, when setting the proportion of people entering the palliative care health state to 0% for both arms, assuming 10,000 patients, the company's model[37] predicts that [redacted] patients in the fedratinib arm enter this health state, which should not be possible.

Finally, following examination of the correspondence between the description of the model reported in the CS[1] and the company's executable model,[37] two additional programming errors were identified, and corrected by the company at the clarification stage:[18]

- When running the model with 10,000 patients, excluding transitions to the AML health state, [redacted] of responders and [redacted] of non-responders were predicted to discontinue treatment before Week 24. The ERG sought clarification[18] from the company on this issue (see clarification response,[18] question B5). In their response the company amended the economic model to correct for this error. This error did not affect the CS base-case results[1] and was only visible when sampling a large number of patients.
- In the CS economic model[37] duration of response was initially applied from Week 24, despite being calculated from time to response. Clarification was sought from the company (see clarification response,[18] question B15) and this was subsequently corrected by the company. The ERG notes that the impact on the ICER was minimal.

4.3.4.3 Concerns regarding health state included in the model

The ERG believes that basing key health states according to the treatment received (on JAK, on BAT) to be appropriate, and in line with TA386.[13] However, as described below, the ERG does not consider the inclusion of AML and palliative care (end of life) to be appropriate, as implemented in the company's model.[1, 37] The ERG also believes that in line with TA386,[13] supportive care should be included as a health state to reflect the period of time prior to death where the disease is no longer controlled with patients receiving only supportive treatments. It should be noted that in the company' model, the palliative care health state reflects inpatient care in the final 8 weeks of life; often referred to as end of life and therefore this is different to supportive care (see CS, Section B.3.3.4.2, page 136).

4.3.4.3.1 Absence of a supportive care health state and subsequent lack of face validity for the scenario assuming worsening in utility.

In TA386,[13] four health states were considered: (1) on JAK, (2) on BAT, (3) on supportive care and (4) death. Patients in the BAT health state were assumed to receive BAT (mostly hydroxyurea) which provides some symptom relief and control of haematological parameters but has limited impact on HRQoL. Following BAT, patients could move to supportive care, consisting of RBC transfusions and

palliative management/monitoring of the disease progression, until death.[13] There is therefore a clear distinction between BAT (e.g., patients are on treatment that provide symptom control) and supportive care (e.g., treatments that are supportive but symptoms are no longer controlled) in TA386,[13] with quality of life in patients entering the supportive care health state worsening with time (to reflect the progressive nature of MF when patients receive supportive care treatments).

The company model[1, 37] does not include a supportive care health state, but instead, the company states (CS, Table 30, page 96)[1] that "*Clinical feedback indicated that BAT and 'supportive care' were equivalent*". Following review of the advisory board notes (both original[27] and final[37]) no reference was identified by the ERG to support this assumption. Clarification[18] was sought from the company to provide a clear reference to this assumption (see clarification response,[18] question B17). In response the company stated that "*The intended interpretation here is that patients who are on BAT in the setting of relapsed/refractory or intolerant to ruxolitinib are on a therapy that can be considered supportive care, in that it does not achieve a trial endpoint*".

The ERG does not consider the response from the company to be satisfactory, and is also inconsistent with the company argument that patients are continued on ruxolitinib (89% of BAT patients in the company's base-case) in the absence of alternative therapy. The ERGs clinical advisors reported that patients may enter a period of time when they will stop treatment prior to death, with the disease progressing, and that this is period is more reflective of supportive care when patients are no longer on treatment that provides some symptom control.

Furthermore, following the company's argument, it is the ERG's view that the assumption for patients initiated on fedratinib to move to a similar BAT health state (without the cost) as the comparator arm to be therefore debatable, and consider that after 2 JAK, patients are more likely to move to supportive care (if JAK are not continued). The ERG notes that in HMRN (2020), of the [redacted] patients who discontinued ruxolitinib, [redacted] did not receive any further treatments.

Given the lack of distinction between the BAT and supportive care health state in the company's model,[1, 37] the CS scenario analysis (to reflect the scenario in TA386[13]) exploring the progressive nature of MF (worsening in HRQoL) applies a reduction in HRQoL to the BAT health state, despite 89% of patients being assumed to receive ruxolitinib and having symptom control. Consequently, the ERG does not consider the scenario as conducted by the company to be valid or relevant. The ERG considers that a full rethinking is required by the company and that supportive care should be included as a health state.

4.3.4.3.2 Concerns with the implementation of AML as a health state

While the ERG considers it to be important to capture transformation to AML, as this is a feature of MF, the ERG does not consider the approach taken by the company to include AML as a health state to be appropriate.

In brief, little detail is first provided in the CS,[1] with details extracted from the economic model.[37] The CS[1] states that it is unclear whether treatment influences the rate of AML transformation. Recognising the uncertainty in this parameter, the ERG considers that assuming the same rate to be reasonable. However, the ERG believes that using data from the long-term COMFORT trial for ruxolitinib[43] (as per the company' notes in the economic model[37]) to be more appropriate compared with data from JAKARTA-2[11] (value actually used in the economic model[37]) as both the follow-up (126.3 vs [redacted] weeks) and the number of patients (146 vs. 97) was greater in the COMFORT trial.[43]

Perhaps more importantly, as previously described in this ERG report, there are some programming errors in the CS. In the company's base-case,[1, 37] survival for patients with AML is re-estimated using a formula described in Appendix L.8 of the CS.[37] The ERG notes that when running the company's base case model[37] for 10,000 patients, the model predicts that [redacted] patients in the fedratinib arm would transform to AML, of whom [redacted] would have a predicted time in AML of less than zero, which is not possible. Patients who are predicted to experience AML, with a time less than 8 weeks are also assumed to move directly to palliative care and therefore have a time in health state equal to zero. The time in the AML health state is estimated to be zero for [redacted] patients ([redacted]%) with AML (while the ERG understands this was intentional, this highlight some conceptual issues to the ERG).

Recalculating OS also led to different OS prediction compared with the initial parametric fit. If the option to re-calculate survival is disabled (which is rightly not presented by the company), the time predicted in the AML health state is over-inflated to a mean time of [redacted] weeks (median: [redacted] weeks), which does not align with the survival estimate from Mesa et al (2005)[42] presented by the company in Appendix L.8[37] (median survival of less than 12 weeks). Approximately [redacted]% of patients with AML (n=[redacted]) have a predicted time in this health state of more than 10 years.

The company therefore had to recalculate survival, but this is not done correctly and a large number of individuals have a time in AML equal to zero. Consequently, the ERG does not consider that the company has implemented the AML health state correctly as it generates negative values, a large proportion of patients with AML with a time in health state equal to zero and the estimated survival (after recalculation) do not match the initial OS distribution (as it is adjusted)."

4.3.4.3.3 Inconsistent assumptions for palliative care between arms

In the CS, patients are assumed to enter the palliative care health state in the last 8 weeks of life. The ERG does not consider the approach taken by the company to be appropriate, as patients on fedratinib (for whom the time to death is the same or less than the time to treatment discontinuation) cannot enter this health state. Therefore, whilst almost all patients on BAT enter this health state, this is not the case for fedratinib. Indeed, it is predicted that [redacted]% of patients on BAT enter this health state, compared with [redacted]% of patients initiating fedratinib. This is because of the sampling approach used in the CS. The ERG notes that this limitation is recognised by the company.

4.3.4.4 Concerns regarding the lack of consistency between assumptions made for the comparator and the intervention arm

A key assumption in the company's model[1, 37] is that patients in the comparator arm (BAT consisting mostly of ruxolitinib) remain on treatment for life, until death, unless patients enter the AML or the palliative care health state for the last 8 weeks of life.

This is primarily justified in the CS[1] (CS,[1] page 132, Section B.3.3.4.1) by the lack of alternative treatment options. This assumption therefore attempts to reflect UK clinical practice. This is broadly in line with the ERG's clinical expert's opinion that patients would typically derive some benefit from ruxolitinib while not achieving a full response, and treatment would be continued as long as possible in the absence of alternative treatments. The ERG's clinical advisors, however; noted that whilst patients would be maintained on treatment (ruxolitinib) for as long as possible, patients may eventually discontinue either because of AEs or when patients are approaching end of life, although the timing is uncertain.

In contrast, in the company's model,[1, 37] fedratinib is assumed to be given and stopped (as observed in JAKARTA-2[11]) based on the TTD, with patients subsequently receiving BAT (consisting mostly hydroxyurea [HU]) for the remainder of their life.

The ERG, supported by its clinical advisors, believes the approach employed by the company to be inconsistent, and that patients initiated on fedratinib (the majority [89%] assumed to switch from suboptimal ruxolitinib) would also remain on treatment (suboptimal fedratinib) until end of life, in the absence of alternative targeted treatments (as justified by the company for ruxolitinib).

The ERG further notes that for the comparator arm, the CS economic model[37] predicts that patients spend [redacted] years in the BAT health state (comprising of mostly ruxolitinib treatment – and therefore incurring high costs in this health state), but spend a similar amount of time in the BAT health state ([redacted] years) following fedratinib discontinuation, incurring low costs (HU) but the same benefit.

The ERG sought clarification[18] from the company on why different assumptions are used between the comparator and fedratinib arm (see clarification response,[18] question B6). In response, the company stated that "*In the absence of data to suggest that fedratinib would continue after patients have lost response, patients were modelled to move to BAT excluding a JAK inhibitor. This is also consistent to the approach taken in TA386*." The ERG does not consider this argument from the company to be satisfactory. An analysis was further requested by the company to reflect that patients initiating fedratinib would remain on treatment despite suboptimal response because of the absence of alternative targeted therapy (as per the assumption assumed for BAT/ruxolitinib). In its response[18] the company provided an analysis where fedratinib is continued in responders only. The company provided the following justification "*If the assumption is made that patients can continue fedratinib beyond the current time-to-discontinuation, it may only be reasonable to assume this occurs in patients who initially responded. These are the patients who continue suboptimal ruxolitinib. It is expected that non-responders would discontinue fedratinib according to the time-to-discontinuation curves (or sooner with the stopping rule).*" ICERs increased from £13,905 to £62,014 per QALY gained assuming all responders (100%) to continue treatment for life. The ERG notes that the company generated ICER using the stopping rule for non-responders, and therefore all non-responders are assumed to stop treatment at Week 24, but assumed to have no change in survival; which the ERG does not consider to be appropriate as described in Section 4.3.4.10. The ERG notes that when non-responders are assumed to continue treatment as per TTD (in line with the description in the company response), the ICER increases to £77,042 per QALYs gained.

The ERG believes the justification provided by the company to be inconsistent with the key CS[1] argument that patients are kept on ruxolitinib while achieving a suboptimal response in the absence of alternative therapy. Should fedratinib be available, no alternative treatment is available for patients who are non-responders at 24 weeks as ruxolitinib would have been discontinued. Therefore, following the company's logic, the ERG believes that patients would remain on fedratinib. The ERG further notes that this is inconsistent with the assumption that 89% of patients in the comparator arm are continued on ruxolitinib when 55% of the JAKARTA-2 population is considered to be refractory to ruxolitinib.

While the ERG recognises challenges in defining how long people remain on treatment, the ERG believes that assumptions for both arms need to be consistent with each other. The ERG also believes that the model needs to reflect how treatment will be used in practice (as done for BAT), rather than as given in a trial setting (as done for fedratinib).

4.3.4.5 Concerns regarding the OS survival function selection process for patients initiated on fedratinib

OS for fedratinib and comparator is taken from two separate sources and is not adjusted for differences in patients characteristics, so the final comparison is a naïve indirect comparison. Concerns are discussed in Section 4.3.4.6.

In the CS, OS in patients initiated on fedratinib is taken directly from JAKARTA-2,[11] with a parametric model fitted to the trial data. The ERG notes that data from JAKARTA-2 are relatively immature (because of the clinical hold in the trial) and; therefore, fitting a model to the trial data is likely to be associated with considerable structural and parameter uncertainty.

OS for fedratinib has a large impact on the ICER in the economic model as patients are assumed to stop treatment early but could experience large benefit following treatment discontinuation (as OS and TTD are modelled independently of each other).

The ERG considers the description of the extrapolation method in the CS to be unclear. As part of the factual accuracy process, the company acknowledged the lack of clarity and provided the following details. The response from the company is reproduced here: "*It is appreciated that the CS and advisory board report potentially lacks clarity on the issue of the OS survival function selection process, and this has led to factual inaccuracies in the ERG report. For clarity, the advisory board was approached with the OS KM for the ITT population, which is why the advisory board report only includes the ITT KM and extrapolations. During the advisory board, the company was informed that ITT was not appropriate for the UK population, and that the intermediate-1 patients (16%) should be excluded. In figure 2 of the ad-board report, the clinicians agreed that for the ITT curves, that exponential and Weibull distributions were reasonable, and was why Weibull was selected as the ITT base case. The clinicians gave the 'consensus values' on what the intermediate-2/high-risk population would be expected to look like. The clinicians agreed that for this population they would expect a curve of a similar shape to the preferred ITT curves selected, but poorer outcomes to account for the higher risk group. The closest curve to the consensus values was the Gompertz. Following the advice given at the advisory board, the intermediate-2/high-risk subgroup was assessed and extrapolated. It was found that the Gompertz curve was the closest curve to the consensus estimates, therefore it was used in the base case*."

The ERG is not able to provide a full critique, given that this information was submitted during the factual accuracy process and not the original submission. Despite additional information being provided by the company during factual accuracy check, the ERG remains unclear about a number of aspects; and therefore, the critique included in this section reflects the ERG's interpretation based on the CS, clarification question and factual accuracy check. It is possible that some of these descriptions may be

factually incorrect due to the lack of clarity on the company's part (as was the case in the original ERG description).

The ERG believes that the process used to generate a preferred survival function has some limitations:

- It is difficult for clinicians to distinguish between models on the survival function scale, although different models have different underlying hazard functions.
- In response to clarification questions A18, A21 and A22, the company wrote that the clinical study team did not consider the shape of the expected lifetime hazard hazard functions a priori and the a priori expectation of the clinicians were not collected.
- In response to clarification question A22, the company presented empirical hazard functions, although it is not clear how these were used in the selection of a preferred model for the data.
- In response to clarification questions A18, A21 and A22, the company wrote that "clinical advice was sought", although it is unclear whether this was from the same clinicians involved in the advisory board meeting or whether they were shown the empirical hazard functions.
- Strictly, a Gompertz distribution, which was chosen as the base case, has a monotonically increasing hazard function, although it is not clear whether this is what clinicians intended. The ERG further notes that the parameter estimates that the company generated for the Gompertz distribution includes negative values. When the shape parameter is negative, a proportion of patients are estimated to be immortal, it implies that the mode of the survival times can be negative or zero and that the hazard of an event can be negative.
- Survival functions depend on the characteristics of the defined patient population. It is not clear whether the population discussed in the context of the JAKRATA-2 study is the same as the target patient population. Hence, the uncertain survival functions that were discussed may not represent the survival function of interest.
- the process retrospectively makes use of experts' opinions with knowledge of the sample data, and it is possible that the sample data is effectively being used twice
- No allowance is made of uncertainty about the experts' estimates of the proportion of patients alive.
- Although the current process requires a preferred survival function, there is a presumption that the clinicians are able to state with certainty which is the true model for the data.

New details provided by the company raised further questions. While the company states at factual accuracy check that "*The clinicians gave the 'consensus values' on what the intermediate-2/high-risk population would be expected to look like*", it was the ERG's understanding that the 'consensus values' were defined prior to the advisory board (as the questionnaire was sent prior to the advisory board and values are included in the ad board notes in the Figure 2 shown to experts), and therefore prior to the

company deciding to focus on the intermediate-2/high risk group. The CS states on page 124 that "*Prior to the UK advisory board held for fedratinib, clinician attendees (N=7) were asked to consider and provide their expectations of survival in the post ruxolitinib population for those treated with BAT and those treated with fedratinib*." There is no mention in the advisory board notes that clinicians were asked to amend their estimate during the ad board. The ERG also looked at the wording in the pre-read material/questionnaire sent to clinical expert (when asked to provide estimate of survival) prior to the advisory board and note that the population was defined as "*adults with disease-related splenomegaly or symptoms caused by primary myelofibrosis, PPV myelofibrosis or PET myelofibrosis who have been previously treated with ruxolitinib*" and "*currently, the economic model uses intention-to-treat (ITT) data from the full JAKARTA-2 trial population (N= 97)…*" It is therefore not entirely clear to the ERG whether clinicians gave their expectation for the ITT population or a different group, as values were given prior to the advisory board (prior a decision was made to focus on the intermediate-2/high risk). However, this is not entirely clear and this could be a misunderstanding by the ERG due to the lack of clarity on the company's part in the description of the process in the CS, advisory board notes and factual accuracy check.

More importantly, the ERG further notes that quantifying experts' beliefs was not done using formal elicitation methods. Pooling of estimates across experts was done using simple averaging rather than behavioural aggregation, which is the approach preferred by the ERG, and no attempt was made to quantify uncertainty in the experts' estimates. The experts had a range of opinions (Figure 22) which reflects the extent of the parameter and structural uncertainty. Finally, the company appears to have selected the preferred model after the advisory board with no further clinical validation (ERG interpretation of wording at the factual accuracy check). Although again; this could be could be a misunderstanding by the the ERG due to the lack of clarity from the company's part from its response at factual accuracy check.

Overall, it is the ERG's view that the process used by the company has several limitations, although the ERG is unable to provide a full critique given the lack of clarity in the CS, clarification questions and factual accuracy process.

The exploratory analysis conducted by the ERG use the generalised gamma distribution due to the lack of clarity and because this was closer to the survival function selected by the experts during the advisory board meeting (when shown the fit to the ITT population with the correct population in mind) rather than the model selection by the company post-advisory board.

Figure 22: Prediction for OS for fedratinib in JAKARTA-2 using the Gompertz and generalised gamma distributions predicted by the model against estimate for the 5 clinical expert prior to the advisory board

In addition, given that OS for fedratinib and the comparator are modelled independently from each other, the CS[1] notes that selected survival functions are predicted to cross at around 6 years and that it was not expected that the survival of BAT patients would exceed that of fedratinib patients at any point (CS,[1] page 130). A constraint was therefore added in the CS for the hazard of death to follow the BAT arm at the point of crossing. The ERG believes that it is a strong assertion to believe that "*it was not expected that the survival of BAT patients would exceed that of fedratinib patients at any point*" (CS, Page 130) because this ignores uncertainty in the true effect of fedratinib and the relative hazard of death in patients surviving beyond six years. It is unclear why it was not assumed instead that the hazard of death for the BAT arm follows the hazard for fedratinib at the point of crossing, given the plateau predicted for the comparator arm.

The ERG further notes that OS is taken from the subset of patients with intermediate-2/high risk from the original JAKARTA-2 analysis,[11] not according to the stringent definition (re-analysis) of JAKARTA-2[24]. These data were not presented in the CS; therefore, it is unclear to the ERG if those patients have a different prognosis.[18]

4.3.4.6 Concerns regarding evidence used for OS for the comparator and resulting predicted survival

The ERG recognises challenges associated with single arm studies and associated uncertainty with this study design. However, while it is possible for fedratinib to be associated with a survival gain, it is of the ERG's view that evidence is currently lacking, and that no robust or convincing evidence has been presented by the company to support the assumption that fedratinib would be associated with a difference in survival compared with BAT (comprising mostly of ruxolitinib).

In the absence of a head-to-head trial, the CS[1] (Section B.3.3.3) reports results from a systematic review of the literature to identify sources for OS post-ruxolitinib discontinuation. The company identified 13 studies, of which four are subsequently included in the economic model (CS, page 123), with findings from Schain et al (2019)[38] conducted in Norway/Sweden subsequently used in the CS base-case. Issues of comparability of this study compared with JAKARTA-2, and other studies identified in the systematic review are described in Section 4.3.4.6.1. OS for fedratinib and comparator is taken from two separate sources and is not adjusted for differences in patients characteristics, so the final comparison is a naïve indirect comparison despite important differences between population. In particular, it is the ERG's view that the studies are not directly comparable and that the population in Schain et al (2019) does not reflect the population entering the model (consisting mostly of patients that are continued on ruxolitinib).

The CS[1] recognises (CS,[1] page 18 and page 169) that there is limited evidence for OS in people continuing ruxolitinib, with the exception of data from SIMPLIFY-2,[10] and states that OS used in the model for BAT (from Schain et al[38]) is consistent with estimate from SIMPLIFY-2. The ERG generally agrees that OS from SIMPLIFY-2[10] provides evidence more reflective of the population entering the economic model, although not without limitations (as the population between JAKARTA-2 and SIMPLIFY-2 is different according to the ERG's interpretation of their inclusion criteria) and re-iterate caution in comparing studies naively. The ERG further recognises that patients in SIMPLIFY-2 were allowed to cross-over after 24 weeks, and therefore, information is limited to OS before this time point. However, the ERG does not consider predictions from the company to align with SIMPLIFY-2. OS from SIMPLIFY-2 is also more consistent with that in JAKARTA-2 (when studies are compared naively as done by the company). This issue is described in Section 4.3.4.6.2. The ERG further notes that in response to clarification question A12, the company asserted that an MAIC could not be conducted with respect to OS, despite the KM being available and a MAIC conducted for response rate.

The ERG further believes that not all sources of evidence have been fully explored and discussed by the company. In particular, evidence from the COMFORT-trials (RCT for ruxolitinib in JAK-naive) could be used to inform overall survival in patients that are refractory (and still on ruxolitinib) and survival in patients who had an initial response, and discontinued ruxolitinib due to either relapse or

intolerance. While the ERG recognises the difficulty in comparing between studies, and associated limitations (and exploratory nature), it is the ERG's view that evidence from the COMFORT-trials provide a more informative comparison against fedratinib in JAKARTA-2 compared with Schain et al.[38] While the ERG caution with comparing studies naively (without adjustment), a naïve comparison of JAKARTA-2 and COMFORT-2 do not support the assumption of difference in survival between fedratinib and BAT. This is described in Section 4.3.4.6.3.

The company also presents results from a pre-advisory board and supportive evidence from JAKARTA[14, 25] (fedratinib-placebo controlled trial in ruxolitinib naïve patients) to support the assumption of a possible difference in survival between fedratinib and BAT. The ERG does not consider evidence from JAKARTA to necessarily support a difference in survival. A critique is presented in Section 4.3.4.6.3 and Section 4.3.4.6.5 respectively.

The ERG notes that because of the inconsistencies in assumptions between treatment arms described in Section 4.3.4.4 (treatment for life for the comparator arm, but discontinuation allowed for fedratinib), a less favourable survival gain leads to improved cost-effectiveness in the CS model.[37]

4.3.4.6.1 Summary of concerns with comparing studies identified in the systematic review with OS from JAKARTA-2

The ERG has a number of concerns with the approach taken by the company to compare "naively" studies identified in the systematic review with OS in JAKARTA-2.[11] The ERG notes that survival functions for a given treatment will vary across studies depending on the inclusion/exclusion criteria of the studies and the actual mix of patients in a study. It was not possible to adjust for differences in patient characteristics between the patients treated with BAT in the four studies and those treated with fedratinib in JAKARTA-2. Hence, the ERG does not believe that a direct comparison of the fedratinib survival function with the BAT survival functions provides robust evidence of treatment benefit in the target population. The ERG further notes:

- First, the ERG notes that all studies identified in the CS[1] systematic review are conducted post-ruxolitinib discontinuation, and therefore reflect OS at the point at which ruxolitinib was discontinued, rather than at the point where patients become resistant/intolerant to ruxolitinib and "should" discontinue ruxolitinib as per the population entering the CS model. Therefore, the ERG believes that this population is further along the patient's treatment pathway than the population in JAKARTA-2[11] or entering the population in the economic model.[37]
- Secondly, patients in studies identified and subsequently used in the CS[1] no longer received ruxolitinib. The CS,[1] (referencing discussion with the company's clinical experts) states that (a) those who continue suboptimal ruxolitinib are expected to have a similar survival as those

observed in the literature (CS,[1] page 73) and (b) the proportion of ruxolitinib in BAT was not expected to have a significant impact on overall survival (CS,[1] page 170). When cross-referencing the CS[1] with the minutes of the advisory board meeting held by the company in April 2020 (originally sent to the ERG[27]), the ERG notes the following statement "*If patients are on ruxolitinib will it impact on OS? Answer - yes – even if patients are cytopenic & receive sub-optimal dose of ruxolitinib you do tend to see a survival benefit*". In response to clarification[18] regarding this inconsistency (see clarification response,[18] B9), the company referred to the final advisory board notes.[37] The ERG notes that in the final advisory board notes,[37] there is a small variation in language "*Question - If patients are continued ruxolitinib will it have an impact on OS? Answer – Potentially, even if patients are cytopenic & receive a sub-optimal dose of ruxolitinib you may see a very small survival benefit*". The ERG is not able to explain this discrepancy, but the ERG's clinical advisors report that patients continuing on ruxolitinib would derive some OS benefits.

- Thirdly, the CS appears to ignore the limited number of studies that report data on OS in patients receiving ruxolitinib following ruxolitinib failure, despite assuming that 89% of patients in BAT receive ruxolitinib. The CS[1, 37] includes four studies in the economic model (Schain et al, 2019;[38] TA386,[13] Kuykendall, 2017[39] and Palandri et al, 2019[40]). This is justified in the CS by the availability of the KM survival functions being available (CS,[1] page 123). The ERG notes that Mehra et al (2016)[59] was identified by the company in the SLR and reports OS in patients who had received frontline ruxolitinib, and received second line therapy (separated between patients receiving ruxolitinib or other conventional therapies). Evidence from this study was not considered in the company's economic model and no justification for its exclusion was provided in the CS.[1] The exclusion and lack of justification of this study (when evidence on OS in people continuing to receive ruxolitinib is already limited) raises significant concerns with the approach taken by the company to both identify and select evidence. The ERG sought clarification[18] from the company (see clarification response,[18] question B7). In their response the company stated that "*…the baseline characteristics for the patients who received ruxolitinib as a front-line therapy was not; primarily, the proportion of patients with intermediate-2 or high-risk classification was unknown. As such, the data is likely to contain patients that would not be eligible for fedratinib in the UK. Given that there were other studies, that were included as options within the model, which did report relevant baseline characteristics in some form, the Mehra study was not considered relevant as the outcomes could not be interpreted alongside any information on MF classification*". While the ERG agrees that baseline characteristics (risk groups) are not available in this study and this is an uncertainty, the ERG considers that excluding studies reporting outcomes in people receiving ruxolitinib (representing 89% of the comparator arm), when evidence is already limited, to be an important omission. In particular, the ERG notes that Kuykendall, 2017[39] and Palandri et al, 2019[40] are

used in the CS scenario analysis despite including some patients classified as intermediate-1. The ERG further notes that whilst the baseline characteristics are unknown, evidence from the Mehra study[59] study could be used in some way to estimate the relative effect of ruxolitinib vs. non-ruxolitinib therapies on OS and be applied in the model (if there is an appropriate BAT OS curve excluding ruxolitinib).

- The CS uses evidence from Palandri et al (2019)[40] (data from novel agents) as a proxy for OS for fedratinib in a scenario analysis and states in Appendix L6[37] that Palandri et al. (2019)[40] "*provides supportive evidence of prolonged survival post-ruxolitinib with 'novel agents' such as fedratinib when compared to 'conventional agents'*". The ERG notes that in Palandri et al. (2019)[40] more than a third of patients (35%; n= 11/31) received ruxolitinib compared with only 3% (n=1/31) receiving fedratinib. The company was asked (see clarification response,[18] B20) why it believed appropriate to use this source of evidence for fedratinib, but not as a proxy for BAT (mostly ruxolitinib). In response, the company commented that this study included intermediate-1 patients, excluded patients in blast phase and that it included other investigational therapies. While the ERG does not believe evidence from Palandri et al. (2019)[40] to be relevant as this included other agents, the ERG finds it illogical that the company uses this study for fedratinib in their scenario analysis and to justify a difference in survival, but considers this study as inappropriate for BAT (mostly ruxolitinib) when more patients received ruxolitinib in this study.
- OS from Schain et al (2019)[38] is used in the base-case for BAT. There was a large initial drop in survival in Schain et al (2019).[38] This was highlighted by clinical advisors to the company in the advisory board notes.[27] In contrast, in the CS OS in patients initiating fedratinib is taken from JAKARTA-2[11] and patients had to have a life expectancy of more than 6 months to enter the trial (inclusion criteria). Therefore there are concerns with comparing OS from a trial (JAKARTA-2[11]) which has strict inclusion criteria, compared with the survival from an observational study. The company was asked (see clarification response,[18] B9) to comment on the comparability between studies (Schain et al, 2019[38] and JAKARTA-2[11] given the differences in inclusion criteria. In its response the company acknowledged: "*that this may be a source of potential bias in the results. This exclusion criterion was a not an objective criterion and given the lack of information on the number excluded by this criterion it would have been difficult to adjust for the comparison to BAT observational data*." While the ERG recognises the challenges to adjust for this, the response from the company highlights issues with naively comparing OS from a clinical trial with strict inclusion criteria against an observational study. In particular, the ERG wishes to highlight that at Week 4, the difference in survival between fedratinib in JAKARTA-2[11] and Schain et al, 2019[38] is already large (98.3% vs. 88.9%); the ERG questions whether such effect on survival (if any) would materialise that soon.

- Schain et al (2019)[38] included mostly patients with primary MF (PMF) [99.5%], whilst JAKARTA-2 included patients with PMF [redacted], post-ET [redacted] and post PV [redacted], with patients in the latter two subgroups expected to have a better survival compared with PMF. To verify this, the KM survival functions for each subgroup from JAKARTA-2[11] were requested as part of clarification process[18] (see clarification response,[18] B10). As expected, visually, OS for PMF is shorter than other MF type (see Figure 23). It should be noted that the company highlight that the differences in OS were not significant between subgroup and that there was a large degree of uncertainty. The ERG agrees that there is uncertainty, but notes that absence of evidence that survival functions are different is not the same as evidence that they are the same and that the survival functions are clearly visually different. This further suggest that that OS from Schain et al (2019)[38] is not directly comparable to the whole JAKARTA-2 population.

Figure 23: Overall survival for PMF and other subtype for the intermediate-2/high risk population

- Patients included in Schain et al (2019) were also significantly older compared with JAKARTA-2. The median age at diagnosis was 70 years in people who discontinued ruxolitinib. In JAKARTA-2, the median age at entry was 67 (38-83), with patients entering the trial at a median of 4.1 years since MF was diagnosed.
- The company fitted parametric models to the BAT data from Schain et al (2019) and selected the Weibull survival function. The CS stated (page 125) that "*the group indicated that the exponential and Weibull were most relevant and representative of UK patients. The Weibull curve was selected in the base case as it provided a better statistical fit to the data*". The ERG notes the following statement in the advisory board notes that "*Exponential & Weibull extrapolations are the most relevant to the dataset (see figure 1). Exponential may better*

represent the survival to UK patients." While the Weibull distribution provide better statistical fit to the data, the exponential distribution (appeared to be preferred by clinical experts at the advisory board) and the Weibull distribution provided very different long -term extrapolation.

- Finally, when describing studies for OS, the CS[1] selected Schain et al (2019)[38] following clinical advice to represent OS, but mentioned later on that the study was deemed inappropriate for a UK setting by its clinical experts, as results were presented solely from Sweden and Norway (CS,[1] page 153). Clarification[18] was sought (see clarification response,[18] B9) on this inconsistency, with the company acknowledged the poor wording and stated that it refers to the proportion of ruxolitinib use only. The ERG has some concern with taking survival from a source, when the treatments received are not deemed reflective of clinical practice. The company, however; justify this from its discussion with clinical expert stating that "*any benefit to OS from receiving ruxolitinib would not be significant*" (see clarification response,[18] B9). The ERG notes that absence of evidence of an effect is not the same as evidence of absence of an effect. As previously observed, the wording in the advisory board notes is also less definitive.

4.3.4.6.2 Predictions from the company not in line with SIMPLIFY-2

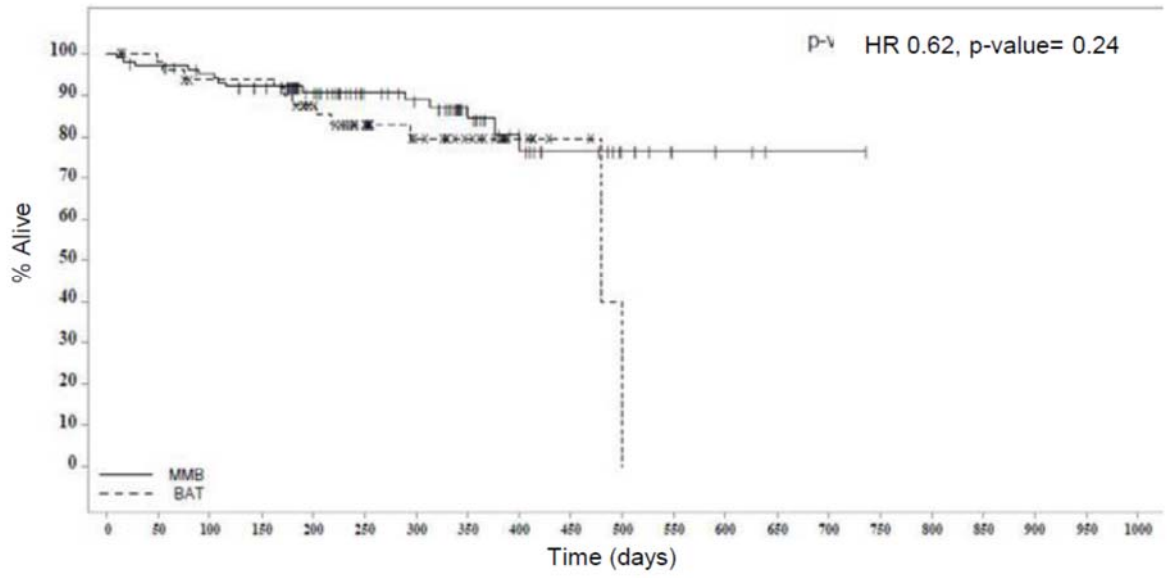
In summary, the ERG does not consider evidence from SIMPLIFY-2 to support the assumption of a difference in survival between fedratinib and BAT.

The CS[1] acknowledges the limited evidence for OS for patients continuing ruxolitinib sub-optimally, and references data from SIMPLIFY-2[10] at 24 weeks for the BAT arm to validate this prediction. The company states that at 24 weeks, in SIMPLIFY-2, 21% of patients died and this is in line with the available BAT KM survival function (CS,[1] page 18, 169).

The ERG notes that 36.3% of patients died at Week 24 in Schain et al (2019).[38] The ERG further notes that in the reference provided,[60] whilst it is stated that "*at the end of the [double blind] DB treatment phase, a smaller proportion of momelotinib patients had died (14% vs. 21%; HR 0.62, p=0.24*)", the KM survival function is also available, and the value from the KM survival function at Week 24 is less than 10%. The ERG sought clarification[18] from the company for why the KM estimates were not used and to clarify the statement about the consistency between the OS estimates from SIMPLIFY-2[10] and Schain et al (2019)[38] (see clarification response,[18] question B7). In its response, the company stated that "*The KM was not interpreted as we relied upon the authors to provide accurate interpretation of their own data*" and that "*The KM is not consistent, the reported value of 21% patients dying at 24 weeks is consistent*". The ERG considers the response from the company to be unsatisfactory. While the ERG recognises that is difficult for the company to explain the discrepancy between the KM estimates and the estimate reported in the reference material (as this is from a competitor trial), reading the estimate

from the KM survival function requires no less interpretation than taking the estimate from a slide. Therefore, both values (the value in the slide and KM) should have been acknowledged by the company (should the company have doubt about which estimate is most accurate). In fact, the ERG believes the estimate reported in the slide (21%) to be more ambiguous (and requiring interpretation) compared with the estimate taken from the KM survival function at 24 weeks (less than 10%). The ERG notes that the statement in the slide refers to survival at the end of the double blind phase, not at 24 weeks. All patients do not all enter the trial at the same time. It is also unclear how this estimate was calculated and whether censoring was considered. Notably, it can be seen from the KM survival function reproduced in Figure **24** that a number of patients were censored in BAT arm for OS. This also allow a direct comparison with the estimate from the KM survival function at 24 weeks in JAKARTA-2.

Figure 24: OS from SIMPLIFY-2 (Image reproduced and available online from Oncology Sierra 2018[60])



Secondly, the ERG considers the statement from the company that survival in Schain et al (2019) is consistent with SIMPLIFY-2 to be inaccurate. The ERG caution against comparing studies naively but notes that the survival difference (difference of 15%) between SIMPLIFY-2[10] (79%) and Schain et al[38] (64%) at 24 weeks (using the estimates reported in the text; not the KM estimates as preferred by the ERG) is larger compared with the difference in survival (difference of [redacted]) between SIMPLIFY-2[10] and fedratinib[11] ([redacted]% for ITT and [redacted]% for intermediate-2/high risk). Following this logic, it can be inferred, that OS from SIMPLIFY-2[10] is more consistent with OS from JAKARTA-2[11] than with Schain et al[38] (2019). Using the estimate from the KM survival function (as preferred by the ERG as less ambiguous), the OS at 24 weeks in SIMPLIFY-2 become even less consistent with Schain et al (2019) and estimate become similar to JAKARTA-2.

Thirdly, while it is challenging to compare the JAKARTA-2[11] and SIMPLIFY-2[10] populations, the ERG believes from the inclusion criteria of SIMPLIFY-2[10] the relapsed and intolerant subgroup in JAKARTA-2 to be more in line with the population included in SIMPLIFY-2, and that refractory patients in JAKARTA-2 are less representative of the population included in SIMPLIFY-2.[10] The ERG requested that the company provide the KM survival functions for each subgroup (see clarification response,[18] question B10). Visually, the survival (KM) for patients that are either relapsed or intolerant (believed to be more in line with JAKARTA-2[11]) (Figure 26) is worse compared with patients classified as refractory (Figure 25). At 24 weeks, the proportion of patients who died was [redacted]%, [redacted]% and [redacted]% for patients with intermediate-2/high risk that were refractory, intolerant and relapsed in JAKARTA-2 respectively, compared with less than 10% in SIMPLIFY-2 using the estimate from the KM survival function at 24 weeks (21% using the value in the slide).

The ERG recognises the difficulty in comparing studies with different characteristics and that more patients with intermediate-1 are included in SIMPLIFY-2[10] compared with JAKARTA-2[11] (31% vs [redacted]%). However, the ERG notes patients included in SIMPLIFY-2[10] were older (69.4 vs. [redacted] years), had a lower platelet count (126.5 vs. [redacted]) and more patients where transfusion dependent (58.7% excluding unknown vs [redacted]%); all of which would affect survival negatively.

Due to challenges in comparing studies naively, the ERG asked (clarification question A12) the company to confirm the feasibility of conducting an unanchored indirect comparison with respect to overall survival (not specific to SIMPLIFY-2). In its response, the company asserted that an MAIC could not be conducted with respect to OS as it need "*OS Kaplan Meier data and report sufficient data on baseline*". The ERG notes that both the KM and baseline characteristics are available in SIMPLIFY-2. In fact, the company conducted a MAIC for response rate, and therefore a MAIC could have been conducted for OS too.

4.3.4.6.3 An exploratory comparison of OS from JAKARTA-2 and COMFORT-trials do no support the assumption of a survival gain

The ERG urges caution in comparing studies naively. However, as this approach is employed by the company to justify a survival difference, the ERG consider relevant to present a naïve comparison of the JAKARTA-2 and COMFORT-trials in the subset in similar populations. A naïve comparison of JAKARTA-2 and COMFORT-trial do not support the assumption of a survival difference between fedratinib and BAT.

The ERG requested (see clarification response,[29] question A3) the company to provide OS according to resistance vs. intolerance and relapsed vs. refractory. Evidence for OS from the COMFORT-trials is

available in patients that are refractory (and still on ruxolitinib treatment – matching the key target population in the CS) and following ruxolitinib discontinuation following an initial response or early discontinuation. It is therefore possible to compare OS from the COMFORT-trials to OS from JAKARTA-2. The ERG notes that some data based on the reason for discontinuation is also available in Palandri et al (2019). However, as justified by the company (see clarification response, question B20), the study included a large number of intermediate-1 risk and therefore was not considered relevant.

- **Naïve comparison of survival in patients that are refractory (e.g no response)**

The ERG's understanding is that patients that are refractory are patients who do not exhibit an initial response to ruxolitinib. Indeed, in the JAKARTA-2 re-analysis, refractory was defined as: ruxolitinib treatment for ≥ 3 months with < 10% SVR or <30% decrease in spleen size from baseline.

Two sources of evidence were identified by the ERG in a similar population (refractory, that is, continued on ruxolitinib and therefore align with the key target population in the CS [89% ruxolitinib]):

- Miller et al (2017) reports OS from the COMFORT-I trial in patients treated with ruxolitinib who had a reduction in spleen length at Week 12 (<25%, 25-50% and ≥50%). All patients were intermediate-2/high risk,
- Palandri et al (2017) report OS from a landmark analysis at 6 months, separated onto whether patients experienced a response or not (IWG-MRT definition, based on spleen length[37]). 84.3% of patients were classified as intermediate-2/high risk (with non-responders more likely to be intermediate-2/high risk).

While the ERG recognises limitations and uncertainties with comparing studies naively, this comparison shows that there are no visual differences in survival between patients that are refractory in the JAKARTA-2 trial ([redacted]9% of the population [n=[redacted]]) and patients treated with ruxolitinib (key target population in the CS) with no spleen response in Miller et al (2017) and Palandri et al (2017). It should be noted that both studies report OS for non-responders at different time-points (12 vs 24 weeks), but also different definitions for non-spleen response (< 25% vs <50% reduction in spleen length). Overall, it is the ERG's view that a naïve comparison of JAKARTA-2 and COMFORT-trials do not suggest that patients refractory to ruxolitinib initiated on fedratinib experience a better survival compared with patients that are remaining on ruxolitinib treatment (assumed to be 89% of BAT arm). However, as previously stated, the ERG urge caution with any naïve comparison.

Figure 25: Comparison of OS from JAKARTA-2, Miller et al, and Palandri et al.

- **Naïve comparison of survival in patients that either relapsed or intolerant**

The ERG understands and acknowledges the company's response to clarification (see clarification response,[18] question B13) on the comparability of SIMPLIFY-2 and JAKARTA-2 that "*there are no internationally recognised criteria for intolerant or resistance, which means that assignment to these patient groups is open to interpretation and there could be overlap. Clinical advice indicates that there can sometimes be difficulty in defining intolerance, such that patients can have a mixture of both intolerance and resistance*".

In JAKARTA-2 re-analysis, relapse was defined as ruxolitinib treatment for ≥ 3 months with regrowth, defined as <10% SVR or < 30% decrease in spleen size from baseline, following an initial response while intolerance was defined as ruxolitinib treatment for ≥28 days complicated by the development of RBC transfusion requirement (≥ 2 units per month for 2 months); or Grade ≥ 3 thrombocytopenia, anaemia, haematoma and/or haemorrhage while receiving ruxolitinib.

As identified (and used in scenario analysis for BAT OS) by the company, OS following ruxolitinib discontinuation in the COMFORT-II study is available in early discontinuers and spleen responders (n=39). All patients were intermediate-2/high risk.

Again, recognising limitations and uncertainties with comparing studies naively, the ERG exploratory analysis shows when comparing JAKARTA-2 and COMFORT-II trial, that there are no visual differences in survival between patients that are either relapsed or intolerant in the JAKARTA-2 trial and OS following ruxolitinib discontinuation in early discontinuers and spleen responders from the COMFORT-II trial. It should be noted that this is the survival at the point of discontinuation and patients were no longer on ruxolitinib, and therefore, OS from the COMFORT-II trial may be an underestimate. It is therefore the ERG's view that a naive comparison of JAKARTA-2 and COMFORT-trial do not support the assumption of a difference in survival for these subgroups either.

Figure 26: Comparison of OS from JAKARTA-2 and COMFORT-II

[redacted]

4.3.4.6.4 Inconclusive evidence from JAKARTA

Supportive evidence from JAKARTA,[14, 25] a placebo controlled trial for fedratinib in ruxolitinib-naive patients were also presented in the CS[1] to support the plausibility for an improvement in survival in patients previously treated with ruxolitinib. An OS HR of [redacted] (95% CI: [redacted]; p = [redacted]) is reported in the CS[1] (CS,[1] page 75) for fedratinib 400 mg versus placebo.

The ERG notes concerns about the generalisability of the HR to patients previously treated with ruxolitinib; which are recognised in the CS[1]; but also that JAKARTA[14, 25] is a placebo-controlled trial (as stated by the company); and therefore the relative treatment effect is not against BAT (as defined in the economic model). Perhaps more importantly, the ERG notes that the HR for fedratinib 500mg vs. placebo in JAKARTA[14, 25] was [redacted]).

The ERG sought clarification from the company, to comment on the observation that there is a dose related effect of survival with 500 mg fedratinib being similar to placebo (see clarification response,[18] question A24). The company considered that to provide an answer would be speculative, but the similarity in death rate between fedratinib 500 mg and placebo could be attributable to baseline characteristics. Notably, the company states that there was "*noticeably greater percentage high risk status in fedratinib 500mgs (■) and placebo (■) arms compared to fedratinib 400mgs (■) arm but it could also be due to other factors*". While the ERG acknowledges that any answer is speculative, the ERG notes that the imbalances in risk groups highlighted by the company between fedratinib 400mg and fedratinib 500mg to explain the differences in HR between the two different dosages, may also explain the favourable HR for fedratinib 400mg against placebo (imbalances in risk group). The ERG further notes that in JAKARTA, a higher proportion of patients in the 400 fedratinib mg arm had a baseline ECOG PS score of 0 (■%) compared with the placebo and fedratinib 500 mg arms (■% and ■%, respectively). Consequently, it is the ERG's view that evidence from JAKARTA (HR in ruxolitinib naïve patients) do not necessarily support the assumption of a survival difference (given the imbalances between treatment arms, and that patients in fedratinib 500mg arm appear to be closer in terms of characteristics to those in the placebo arm and no difference in survival was observed). It is the ERG's view that should the HR from JAKARTA fedratinib 400 mg be considered appropriate, this raises significant concern about the safety when the dose is increased.

4.3.4.6.5 Caution with interpretation of the pre-advisory board estimate

The company stated that prior to an advisory board meeting (CS, Page 124), seven clinicians were asked to provide estimates of the proportion on patients surviving at 1, 2, 5, 10, 15 and 20 years following treatment with fedratinib and BAT. The ERG notes that this was not done using formal elicitation methods, no attempt was made to quantify uncertainty in the experts' estimates and that pooling of estimates across experts was done using simple averaging rather than behavioural aggregation, which is the approach preferred by the ERG. Furthermore, the following question was asked at the meeting "*Based on your clinical experience and the data available, what are your estimates for survival in the post ruxolitinib population*"?. Briefly, the ERG notes (a) the population is clearly set as "post-ruxolitinib" and (b) the lack of mention that almost all patients (89%) are assumed to continue to receive ruxolitinib while achieving a suboptimal response for the comparator arm. Consequently, the ERG urges caution in interpreting the quantities presented in the CS.[1]

4.3.4.6.6 Conceptual inconsistencies in the model

The company also presents data on surrogacy between spleen and survival to support a survival gain. This scenario is described in Section 4.3.4.9 and lacks face validity. But, more importantly, as highlighted in Section 4.3.4.2.4, the CS base-case model itself predicts no differences in OS between responders and non-responders.

4.3.4.6.7 OS from the HMRN

The CS[1] reports data on OS post-ruxolitinib (in people not receiving subsequent ruxolitinib) from UK patients from the HMRN (n=[redacted]). However, the company appears to suggest that this source is not appropriate and that the HMRN data[52] may not be considered representative of UK clinical practice as the uptake of ruxolitinib in the HMRN is low, in particular that the proportion of patients starting ruxolitinib each year decreased in subsequent years despite a positive recommendation (CS,[1] page 14).

While the ERG agrees that the sample size is small, the ERG considers the statement from the company confusing, as uptake would not affect OS observed post-ruxolitinib discontinuation. While the ERG does not believe values from the HMRN to reflect the population assumed in the company's model (e.g., patients continuing suboptimal ruxolitinib at the point where they become intolerant/resistant), using values from the HMRN lead to a significant survival gain in favour of fedratinib, but have the effect to increase the ICER. This is because, as previously discussed, patients in the comparator remain on treatment for life, whilst patients on fedratinib discontinue treatment early. The ERG further notes that in JAKARTA-2, the median duration of exposure to prior ruxolitinib was [redacted] month against [redacted] years in the HMRN. It is the ERG's view that these two points are suggestive that patients were further along the pathway and that the population in the HMRN are perhaps more reflective of a population who discontinued ruxolitinib (following a time on suboptimal ruxolitinib) and therefore could reflect the time in supportive care.

4.3.4.7 Concerns with using response rate and its application in the economic model

The ERG has a number of concerns with the CS estimation and reliability of response rate and subsequent application in the economic model,[37] most of which are described in Section 3.4. The ERG's concerns include (1) comparability of the population included in SIMPLIFY-2[10] and JAKARTA-2[11] despite an attempt from the company to adjust for baseline characteristics, (2) comparability in design which is likely to be more favourable to fedratinib (no washout period in SIMPLIFY-2[10]) and (3) estimation of the treatment effect using difference in absolute risk and its application to baseline risk on the absolute risk scale.

In the company's economic model,[37] in patients initiating BAT, response rate is estimated by subtracting a treatment effect from the response rate used for fedratinib. The ERG does not believe this approach to be appropriate as the treatment effect is applied in the absolute scale and will therefore lead to inconsistencies when the response rate for fedratinib is lower than the estimated treatment effect.

The ERG notes that a constraint is added in the CS economic model to ensure that response for BAT is equal to zero and does not fall below zero in such situations. However, the ERG notes that such constraints should not be required, should this be implemented correctly.

The company also assumes that BAT after fedratinib does not contain ruxolitinib, but the CS uses data for the response rate where 89% of people received ruxolitinib. Response rate for patients initiated on fedratinib entering the BAT health state is therefore not aligned with the costs assumed.

Response rate for spleen is also defined according to spleen volume in the economic model, and the company states (CS, page 100) that patients with a spleen volume response were the same as patients with a response in palpable spleen in JAKARTA-2. The ERG notes that in SIMPLIFY-2, 5.8% of patients had a spleen response based volume, but that the response to spleen length (by palpitation) is 21%. Consequently, while the choice for response using spleen length or volume does not affect the fedratinib arm, the choice for response assessed by spleen length or volume will affect the response rate for the comparator arm (and therefore quality of life). The ERG further notes that in SIMPLIFY-2, TSS was calculated based on the MPN-SAF, and this is could be less favourable to BAT as the MPN-SAF is less specific to MF symptoms. The company does not present TSS in JAKARTA-2 using the MPN-SAF (despite data collected on MPN-SAF); it is therefore unclear whether the same response rate for symptoms would be observed using both instruments. Exploring TSS using the MPN-SAF would have provided a like for like comparison – although the ERG recognises that the MF-SAF is more appropriate in MF.

4.3.4.8 Concerns with using duration of response and application in the economic model

Only a brief critique is presented here as the impact is minimal. DoR is used in the model to determine how long responders experience utility values for responders. The ERG has a number of concerns which are summarised below. First, DoR for spleen response is used, despite response being based on spleen or symptom response. However, this limitation is acknowledged by the company in the absence of data for symptoms. Second, with only two patients who lose response there is insufficient information in the sample data alone to estimate parameters. The ERG is not confident that any of the parameter estimates are meaningful; in particular, the company reports a negative standard deviation for the lognormal distribution. The ERG does not believe that using information criterion (i.e. BIC and AIC) to judge the goodness-of-fit of models to data with only two events is meaningful. Thirdly, an error was identified by the ERG (described in Section 4.3.4.2.5), and raised during clarification stage (see clarification question B15) in that DoR is calculated from time to response, but applied in the model from Week 24. Finally, as previously described for TTD (Section 4.3.4.2.2), the approach to sampling DoR independently to OS and TTD leads to bias. Indeed, as DoR is truncated by TTD and OS, the ERG believes that the resulting DoR would not match what is observed in the trial.

4.3.4.9 Lack of face-validity for the scenario analysis assuming surrogacy between spleen and survival

In addition to its base-case,[1] the company presents results from a scenario analysis whereby OS is estimated as a function of the response rate and a relationship between response and survival (CS,[1] Table 86, page 181). In summary, the ERG believes the scenario analysis presented by the company to estimate OS based on surrogacy to be inappropriate and lacking face validity. Therefore, only a brief assessment is presented in this section.

Briefly, the company conducted a targeted review of studies to assess the relationship between response and overall survival, and identified 12 studies, of which 3 sources are subsequently included in the economic model (Palandri et al, 2017;[61] Vannucchi et al, 2015[62] and Verstovsek et al, 2012[63]). Given the time and resource constraints, the ERG was not able to confirm whether any additional studies were missed or excluded. Overall, the ERG considers that whilst evidence is suggestive that spleen response could be associated with an improvement in survival, (a) the magnitude of benefit is inconsistent between studies and therefore this is highly uncertain and (b) the relationship has been assessed in people treated with ruxolitinib in first-line only, and therefore it is unknown whether the same relationship would apply following ruxolitinib failure.

The ERG further notes that the surrogacy observed in the ruxolitinib studies may also be confounded by the duration patients with a response or not remain on treatment. Patients without a response are more likely to discontinue treatment sooner compared with responders (as justified by the company for fedratinib; see CS page 133). This will therefore influence the relationship between response and survival. It is unclear if the same relationship would be observed if patients were treated for the same duration.

The ERG further notes that the CS[1] states (CS,[1] page 48) that using SVR as a surrogate for OS was validated by clinical experts, referencing to notes from the company's ad-board. When cross-referencing the CS[1] with the minutes of the advisory board meeting held by the company in April 2020 (originally sent to the ERG[27]), the ERG was not able to confirm the company statement. A statement was included in the final advisory board notes sent following clarification "*Would it be reasonable to use the BAT SVR response from SIMPLIFY-2 as a surrogate marker to say that patients who are continued on ruxolitinib would not have a meaningful survival gain that would be reasonable? Yes*".

In addition to the sources available to quantify the relationship between response and OS (which is already uncertain), a number of options are included within the economic model[37] to conceptually link inputs. The ERG believes that, should the analysis be robust and conceptually valid, results should be broadly consistent with each other. The company report results from key scenarios in Table 86 (CS,[1] page 181) and estimate that fedratinib is [redacted] compared with BAT when OS is estimated using surrogacy.

The ERG replicated this analysis using 10,000 patients and plotted the predicted OS for fedratinib estimated from this scenario analysis and OS generated in the base-case (direct Gompertz fit). It can be seen (Figure 27) that predicted OS for fedratinib (red line) for the scenario analysis described in the CS[1] (CS,[1] Table 86, page 181) is inconsistent with both its base-case estimate (blue line), and the trial KM survival function (black line), raising significant doubt of its validity.

Figure 27: Comparison of OS generated using the surrogacy scenario and CS base-case

Assumptions surrounding this scenario were not assessed by the ERG further because of its lack of face validity as shown in Figure 27. However, the ERG re-iterates concerns with the reliability of using response rate as described in Section 4.3.4.7 along with the use of the BAT survival function to represent OS for non-responders (at 24 weeks). The ERG further notes that when conducting this analysis, the company uses response rate defined in terms spleen or symptom response, rather than spleen only, although this could have been amended in the economic model. The ERG considers that a complete rethink is required by the company, should the surrogacy OS scenario be considered for decision-making.

4.3.4.10Lack of face validity for the scenario using the stopping rule.

The CS[1] presents a scenario analysis where patients who do not respond at 24 weeks discontinue treatment, to reflect the stopping rule for ruxolitinib in TA386[13] (CS,[1] page 98). First, it is unclear to the ERG if a stopping rule would apply in clinical practice for fedratinib in patients previously treated with ruxolitinib, in particular, in relation to the argument from the company that patients are continued on treatment in the absence of alternative options. Perhaps more importantly, should the stopping rule be clinically valid, the ERG does not consider the scenario analysis as conducted by the company[1] to be appropriate and is lacking face validity (as the company cut the cost, but not the benefits).

The ERG notes that in TA386,[13] a stopping rule was implemented as this was part of the license for ruxolitinib. The company was asked to clarify (see clarification response,[18] question B1) whether a stopping rule is present in the expected licensing for fedratinib and whether this would be part of clinical practice. The company responded that "*The fedratinib SPC states that Treatment may be continued for as long as patients derive clinical benefit". As this isn't a definitive stopping rule, the stopping rule was not presented in the base-case, in line with NICE guidance. However, it does suggest that UK treatment guidance on discontinuation should be adhered to. It was also confirmed at the advisory board that a stopping rule would be used if patients had not responded at week 24*."

The ERG agrees that presenting results for the stopping rule as a scenario analysis is appropriate as this is not part of the licence. However, the CS[1] (CS,[1] page 98) states that the stopping rule is based on the British Committee for Standards in Haematology (BCSH) guideline for the diagnosis and management of myelofibrosis (2012).[3 64] The ERG notes that the stopping rule was added in the guideline (update in 2015[64]) following approval of ruxolitinib and therefore reflects the licensing for ruxolitinib. Nevertheless, the ERG recognises that such a stopping rule could in theory apply to other JAK in practice. However, the ERG is unclear whether this stopping rule would apply in patients previously treated with ruxolitinib, as these patients are further along their treatment pathway and have no other treatment options. In particular, in relation to the company argument[1] that patients are kept on ruxolitinib in the absence of alternative therapy. Should fedratinib be recommended, the ERG is doubtful that patients who switch to fedratinib be discontinued after week 24 because of the absence of response, leaving them with no other therapeutic options.

While the ERG acknowledges uncertainties, the ERG does not believe the approach taken by the company[1] to be appropriate to capture this stopping rule, as there is a clear disconnect between OS and the proportion of people that discontinue treatment at week 24 (OS is fixed, irrespective of the proportion of patients who discontinue treatment because of the 24 week stopping rule). For instance, OS for fedratinib is the same irrespective if the response rate at 24 weeks is 0% or 100% (therefore OS is unchanged even if all patients stop treatment at 24 weeks).

In the CS submitted economic model,[37] the 24 week stopping rule only affects costs, but key efficacy inputs (OS) remain unchanged. This is different to TA386, where non-responders at 24 weeks discontinued ruxolitinib treatment and were assigned outcomes for people without ruxolitinib. Clarification was sought from the company on why the stopping rule is not expected to affect outcomes, whether this is line with TA386 and provide an analysis where outcomes are affected by the stopping rule (see clarification response,[18] question B1). In its response, the company stated that "w*hen the stopping rule is enabled in the model, incremental QALYs are lower, because patients transition to BAT sooner. Therefore, outcomes are impacted in the scenario analysis*". The ERG does not consider the response from the company to be satisfactory, as the stopping rule only affects the estimation of quality of life, but not OS as in TA386[13]. The ERG considers that a new analysis is required should the stopping rule scenario be considered for decision-making.

4.3.4.11Concerns regarding the company's cost assumptions

The ERG considers the company's approach to costing drugs per pack to be generally appropriate. However, the ERG does not consider the inclusion of wastage for ruxolitinib to be appropriate in line with TA386.[13] While the CS states that wastage was included, in line with the preferred ERG assumptions in TA386 (referencing the ERG pre-ACD), the ERG notes that in the ruxolitinib FAD that the inclusion of wastage was discussed by the committee and clinical experts provided the following statements "*The Committee heard from the clinical experts that the company's assumption of no drug wastage for ruxolitinib reflected drug usage in clinical practice.... The Committee agreed that there was some uncertainty over whether the drug costs for ruxolitinib used in the economic model reflected the drug costs for ruxolitinib in clinical practice, but agreed that the drug costs used were appropriate because they were based on the same trial data on which the effectiveness inputs were based*". The ERG further notes that the ERG preferred base-case following ACD included no wastage as highlighted by the following statement taken from the ERG response (Section 3: ERG's preferred base-case; page 12) to ACD "*Considering the comments from committee and the revised base case, we present a revised base case to which the revised PAS (discount of ***) is applied. This ERG revised base case acknowledges the uncertainty regarding estimating drug wastage and therefore assumes no drug wastage*".

In addition, ruxolitinib (assumed to be given to 89% of patients on BAT for life) is costed based on the distribution of patients with a platelet count < 100 x 10^9/L and ≥ 100 x 10^9/L from the JAKARTA-2[11] trial, with patients with a platelet count < 100 x 10^9/L receiving low dose ruxolitinib (5mg BID) and patients with a platelet count ≥ 100 x 10^9/L receiving other dosages (10mg BID, 15mg BID). While the ERG finds the approach generally appropriate and in line with the cost structure and licensing of ruxolitinib, the ERG notes that there is a mismatch between the distribution from JAKARTA-2[24] and

effectiveness data used for ruxolitinib/BAT (as response rate is taken from SIMPLIFY-2[10]). While the distribution of platelet count (patients with a platelet count < 100 x 10^9/L) is not reported in SIMPLIFY-2,[10] the mean platelet count was 126.5 (SD: 95.9) x 10^9/L for BAT in SIMPLIFY-2 (arm from which efficacy [response rate] is taken from for the comparator arm) vs. [redacted] (SD: [redacted]) x 10^9/L in JAKARTA-2.[11] Although the magnitude of bias is uncertain without access to data from SIMPLIFY-2[10] the ERG believes that the cost for ruxolitinib to be overestimated by the company and not aligned with the platelet count distribution for the BAT arm in SIMPLIFY-2 (from which efficacy data [response rate] is taken from). The proportion of patients receiving ruxolitinib and the cost of ruxolitinib are key drivers in the economic model. The ERG notes that the proportion of patients with a platelet count < 100 x 10^9/L was 58% (median 91 x 10^9/L) in Newberry et al (2017), 45% in Kuykendall, 2017,[39] and 43.5% (mean 163.9 x 10^9/L) in Palandri et al, 2019.[40]

The ERG further notes that dose interruption is not considered. It is unclear whether this is similar between treatments. The ERG recognises that data were not available in SIMPLIFY-2.

4.3.4.12Concerns regarding the selection process for TTD survival function

Following the response to clarification questions (see clarification response,[18] question B14), the ERG is satisfied with the response from the company to assume a uniform distribution for early discontinuers. For responders, the company[1] selected the exponential distribution for TTD and justify this choice stating (CS,[1] page 135) that "*the exponential curve was chosen for its clinical plausibility, as other curves exhibited long-term plateaus*". For non-responders, the company select the Gompertz distribution (CS,[1] page 133) to "*reflect the expected limited time on treatment for non-responders in this population, despite not being the optimal statistical fit over the observed period (Table 52). Some of the other curves predicted long-term plateaus suggesting that non-responder patients would still be receiving fedratinib (if alive) beyond 10 years, which was not clinically appropriate (Figure 23). This choice of curve ensured that time-to-discontinuation was shorter on average for non-responders than responders*."

The ERG has a number of concerns with the company's approach to selection of the TTD distribution. No reference to TTD was identified in the company's advisory board notes[27, 37] and therefore the ERG is not able to confirm the clinical plausibility or how survival functions were selected. In particular, given the chosen modelling approach, where TTD is truncated by OS, selecting survival functions that predict the shortest time on treatment is less of a realistic approach. The ERG notes that for responders, the exponential and Weibull distributions provides relatively similar fits and extrapolation, with the exponential distribution providing the best statistical fit. The ERG considers that given the immaturity of the data, using the exponential distribution is therefore reasonable for the base-case, but considers that the Weibull distribution could also be deemed appropriate.

For non-responders, the ERG believes that the exponential distribution is more appropriate compared with the Gompertz distribution for the following reason (a) the exponential distribution had the best statistical fit, (b) it remains consistent with TTD assumed for responders (e.g., lower) and (c) assuming the rate to discontinuation to be constant is perhaps more realistic when data are immature and there is no information on the long term hazard.

4.3.4.13Concerns regarding the company's HRQoL assumptions

The ERG considers the use of the MF-8D[46] in the base-case to be generally appropriate and in line with TA386[13] given psychometric properties of the EQ-5D[44, 47] in this patient population. The company[1] derives utility values using a mixed effect model. It is unclear to the ERG how the statistical model was selected and whether an alternative model would have provided a better fit to the data. In response to clarification (see clarification response,[18] question B12), the company stated that "*mixed effects models have been fitted for utility values given they are repeated measures data. Alternative models, such as those presented in Alava et al. (2012), are developed primarily to address three key issues in the utility values: floor effects, ceiling effects and multimodal distributions. Whether these models are practically beneficial for the type of utility values in JAKARTA-2 is unclear. Utility values from JAKARTA-2 do not display a multimodal distribution nor is there a mass of observations at 1, histograms of the utility values can be found in Figure 7. Residual diagnostics of the mixed effect models suggest that the residual assumptions of the mixed effect models are reasonable. Consequently, the mixed effect model is used for utility values*." Overall, the ERG is satisfied with the company's response.

The company includes Gender in the regression model used to estimate utility values. It is unclear to the ERG why Gender was included and why Gender was believed to be prognostic for utility values for response. The company was asked to clarify why Gender was included and provide an analysis removing Gender from the regression model (see clarification response,[18] question B12). An analysis was provided by the company removing Gender. The ERG believes that excluding Gender is more appropriate if it is known not to be predictive. The impact on the CS base-case was limited.

An analysis was also requested by the ERG using the EQ-5D (see clarification response,[18] question B11). The company stated the EQ-5D was not collected in JAKARTA-2[11] and discussed the appropriateness of the EQ-5D. While the ERG understands limitations associated with the EQ-5D in MF, the ERG considers that it is possible to map between the EORTC-QLC 30 and EQ-5D and that such analysis should be presented for transparency and completeness, acknowledging the limitations.

The ERG recognises the limited evidence available regarding HRQoL in people previously treated with ruxolitinib. In the company's model[37] utility values are assigned according to the response status and

therefore it is implicitly assumed that responders and non-responders to fedratinib are the same as responders/non-responders to ruxolitinib or other non-ruxolitinib BAT treatments (HU for instance). The ERG believes that it is reasonable to assume utility values for JAK treatments (fedratinib or ruxolitinib) to be broadly similar (given similar effect). However, the ERG does not believe that it is appropriate to use utility estimate from JAK treatments as a proxy for utility values for non-JAK treatments, as the latter group is likely to have worse quality of life as described in TA386.[13] This is also inconsistent with the assumption made for resource use where JAK treatments are assumed to have a different effect on resources compared with non-JAK treatments. The ERG believes the assumption made by the company (to assume utility values to be same between JAK and non-JAK) to be favourable to fedratinib. Indeed, for the comparator arm, BAT is mostly composed of ruxolitinib (89%). However, patients in the fedratinib arm who move onto BAT no longer receive JAK and are treated primarily with HU (low cost but experience a benefit in HRQoL). While the ERG recognises that different assumptions could be made, the ERG does not believe that assuming the same utility value for BAT comprising of mostly ruxolitinib (89%) and BAT comprising of mostly HU (0.0% ruxolitinib) to be appropriate. The ERG notes that in TA386, utility values at baseline was used for BAT (non-JAK treatments) and consider this assumption to be perhaps more plausible. However, the ERG re-iterates that there is considerable uncertainty given the limited evidence base, and that a number of alternative assumptions could be made.

A number of assumptions/adjustments are also made by the company for the AML and palliative care health state utility values as they are taken from different sources. The ERG does not focus on these assumptions as these health states are not implemented correctly.

4.3.4.14Inclusion of adverse event in the model

Adverse events have a small impact on results and it is challenging to compare across studies. Therefore, only a brief critique is presented here. In the base-case, AEs only affect costs.

The ERG considers using the incidence of AE from SIMPLIFY-2[10] for BAT to be reasonable, as this reflects the proportion of ruxolitinib use assumed in the economic model.

It is unclear to the ERG why only AEs included in TA386[13] were considered in the company's model.[37] Furthermore, only grade ≥3 AEs were included. It is possible that grade <3 AEs may be associated with management costs. The company was also requested (see clarification response,[18] question B23) to include the impact of Wernicke's encephalopathy (WE) as an adverse event in the economic model. The company stated that "*WE is not an expected adverse event for patients receiving 400 mg fedratinib. This in line with the JAKARTA-2 data and supported by the fact that thiamine levels are to be monitored for all patients considered for fedratinib prior to and during treatment (as per US PI/ draft SPC).*

Therefore, WE is not considered a relevant AE for this economic model." While the ERG recognises that the incidence of WE may be low with fedratinib 400mg, the ERG notes that the licensing for fedratinib[15] includes a special warning and therefore consider this to be relevant.

The CS base-case includes the impact of AE on quality of life separately. The ERG does not consider this to be appropriate as utility values are taken from JAKARTA-2[11] and already include the effect of adverse events associated with fedratinib. The ERG further notes that decrement in utility values are taken from a range of sources using varying preference based measures, and therefore there are concern with mixing values from the MF-8D[46] with other measures. However, this is a small issue.

The ERG notes that the incidence of AEs for fedratinib is taken from the intermediate-2/high risk subgroup of JAKARTA-2[11] but considers using values for the ITT population to be more appropriate as this relies on a larger sample size. The ERG further notes results from a recent study by Pardanini et al (2020) on the long-term safety of fedratinib in people with intermediate-2/high risk MF. While the Pardanini et al (2020) conclude that fedratinib is well tolerated in patients who remained on treatment for ≥24 cycles, grade ≥3 pneumonia events were observed with long-term treatment with fedratinib; although it should be noted the sample size of this study (n=28) is small.[65]

4.3.4.15 Uncertainty regarding resource use

Given time constraints, and broader conceptual issues described previously, only a brief critique is presented here. The ERG wishes to highlight that evidence on resource use in the model are subject to considerable uncertainty given (a) the absence of direct evidence in people previously treated with ruxolitinib and (b) the unknown impact of fedratinib on resource use (against BAT). Indeed, the CS uses a mix of evidence in people not previously exposed to ruxolitinib or evidence on the impact of ruxolitinib on resource use as a proxy for the effect of fedratinib against BAT (non JAK).

It is not entirely clear from the CS where baseline resource use are taken from, in particular whether data from the HMRN (2020) is used. Following review of the economic model, baseline resource use (BAT without JAK) is taken directly from TA386, and therefore values from the HMRN (2020) are not used to update those used in TA386 (from HRMN 2016). The ERG is unclear whether the company explored if data from the HMRN (2020) could be split between patients not on JAK vs. on JAK (at present data are present for all patients, or those treated with ruxolitinib only). The ERG is also unclear whether the company requested for resource use following ruxolitinib discontinuation and whether this was available.

The company uses data from the HMRN (2020) to estimate the effect of JAK on resource use. It is unclear whether the same effect would be observed with fedratinib. The impact on A&E visits, hospital

nights, outpatient visit and urgent care is also calculated by comparing resource use for all patients (including patients on ruxolitinib) vs. resource use in patients treated with ruxolitinib only from the HMRN (2020). The ERG does not consider this to be appropriate as the all patients category includes a mix of patients treated or not with ruxolitinib. The company also assumes that the effect is constant over time, which is uncertain.

In the absence of evidence, the effect of ruxolitinib on resource use is used a proxy for the effect of JAK treatments (fedratinib or ruxolitinib) against BAT non-JAK treatment. While this is plausible and reasonable for the majority of resource use, this is less clear for the transfusion requirements.

Finally, the CS reports data from JAKARTA as supportive evidence of the impact of fedratinib on resource use. The ERG notes that the comparator within JAKARTA was placebo. While patients may receive supportive medication, treatments received are not reflective of those in BAT assumed in the economic model.

4.3.4.16 Underestimation of uncertainty in model parameters

Given time constraints, and broader conceptual issues described previously, only a brief critique of the PSA is provided here. The ERG has a number of concerns with the PSA conducted by the company.

The ERG notes that not all important parameters are varied in the PSA, including unit costs, baseline resource use and BAT composition. This is despite information about the uncertainty for resource use available in TA386 and BAT composition from SIMPLIFY-2.

An arbitrary 10% SE is assumed for a number of parameters. While the ERG consider this approach to be generally reasonable when there are no information about the distribution for a particular parameter, the arbitrary SE is used despite information available for:

- the proportion of patients with an AEs,
- impact of JAK on resource use from the HMRN (2020),
- Platelet count distribution,
- rate of AML transformation, and
- response rate for fedratinib

The ERG further considers that a beta distribution (instead of a normal distribution used in the CS) should be used for the proportion of patients with an AE, platelet count distribution (proportion over < 10^9/L), AML transformation rate and response rate.

Costs are varied from a normal distribution. The ERG considers that a gamma distribution is generally more appropriate to represent cost distributions.

Parameters in OS survival functions for both the intervention and comparator are varied using multivariate normal distributions. However, because OS is taken from two separate sources and curves are not allowed to cross, this may introduce biases.

The ERG consider the use of multivariate normal distribution to represent uncertainty about utility values to be appropriate.

4.4 Exploratory analyses undertaken by the ERG

As described in Section 4.3, the ERG has a number of concerns with the CS implemented economic model. As described, while some of these issues are debatable and may reflect matters of subjective opinion, others reflect more serious underlying problems regarding the conceptualisation and implementation of the model and the use of evidence to inform the model's parameters. Consequently, the ERG has serious doubts regarding the validity of any results generated using the CS economic model[1, 37] and would advise considerable caution in their interpretation and use in informing decision-making. These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethinking of the model's logic (and how evidence is used). As such, the impact of these issues on the expected cost-effectiveness of fedratinib is not entirely clear.

For transparency and completeness, the ERG undertook some exploratory analysis to explore the impact of some of these issues on the ICER. However, the ERG advises caution, as these analyses were limited by the company's model structure. The ERG still believes that a full conceptual re-thinking of the model is required, rather than trying to fix the current model as submitted to ensure that no other aspects were missed. Analyses are presented to illustrate the potential impact on the ICER when some of these assumptions (given the current structure) are considered together.

The ERG preferred base-case is comprised of six key amendments to the company's models; including the correction of errors, changes to assumptions and changes to the model's logic (when possible, although this was limited by the company structure/approach); these are detailed below. Analysis were undertaken using the deterministic version of the model only (due to time constraint).

The cumulative effect for each change is presented, rather than the impact for an individual change each time.

- **Exploratory analysis 1: disabling of options in the economic model (minor impact on the model results)**

This analysis consider the following:

- Removal of the AML health state because of concerns regarding its implementation (generate negative time) as described in Section 4.3.4.2.5,
- Removal of palliative care health state because of concerns regarding its implementation (biased against the comparator arm) as described in Section 4.3.4.2.5,
- Using DoR calculated from 24 weeks, rather than time to response (as acknowledged by the company in clarification response) as described in Section 4.3.4.2.5,
- Removal of wastage for ruxolitinib, in line with the committee and ERG preferred base-case in TA386 as described in Section 4.3.4.11,
- Removal of gender from the utility regression model, in the absence of clear clinical rationale as described in Section 4.3.4.13
- Assume thiamine supplementation until treatment discontinuation and 250 mg per day rather than 200 mg.

- **Exploratory analysis 2: ERG's preferred choice for parametric extrapolation**

This analysis consider the following:

- Using the exponential distribution for TTD for non-responders. As described in Section 4.3.4.12, the ERG considers the exponential distribution for TTD for non-responders to be more plausible compared with the Gompertz distribution assumed by the company. The ERG believes that the exponential distribution should be used for TTD for non-responders, based on statistical fit, consistency with responders, and that little data inform this parameter.
- As described in Section 4.3.4.5, the ERG believes the generalised gamma distribution for OS for fedratinib to be more consistent with the advisory board notes (where clinical experts to the company deemed the Gompertz distribution to be most plausible for the ITT population) compared with the Gompertz distribution.

- **Exploratory analysis 3: No difference in survival**

As described in Section 4.3.4.6, the ERG had a number concerns with the approach taken by the company to estimate OS for the comparator (and associated difference in survival). While the ERG urge caution with naïve comparison, based on a review of the evidence provided by the company but also identified by the ERG (SIMPLIFY-2, JAKARTA, COMFORT-trials), it is of the ERG's view that no robust evidence has been presented to support the assumption of a difference in survival between fedratinib and BAT. While the ERG recognises that an OS gain, could be possible; this has been proven yet.

- **Exploratory analysis 4: Adjustment for TTD to ensure predictions match input**

As described in Section 4.3.4.2, the ERG had a number of concerns with the company's modelling approach and showed that the predicted TTD was underestimated compared with the true estimate. It is challenging to amend the model to correctly account for this given the modelling approach (independent modelling of OS and TTD). Consequently, an approximation was made for this analysis. The parameters for the exponential distribution (for both responders and non-responders) was varied using a hazard ratio so that the predicted mean time on treatment becomes closer to the true mean time on treatment as shown below in Figure 28. A HR of 0.43 and 0.59 was derived using trial-error (assuming the exponential distribution for TTD for both responders and non-responders and OS to follow a generalised gamma distribution). While the fit to the data is not visually optimal (the predicted TTD is over-predicted at the beginning and under-predicted at the end), the generated mean TTD is close to the one should the mean estimated from the exponential directly. It should be noted that this cannot be avoided given the constraint imposed by the modelling approach where TTD is truncated by OS. In summary, following adjustment, the model predicted a mean TTD of [redacted] weeks and [redacted] weeks for responders and non-responders versus a true mean TTD (using the direct curve) of [redacted] weeks and [redacted] weeks, respectively.

Figure 28: Prediction for TTD following adjustment

[redacted]

- **Exploratory analysis 5: Assumption that 88.5% of fedratinib responders continue treatment**

As described in Section 4.3.4.4, the company assumes that patients on fedratinib are allowed to discontinue treatment as per TTD, whilst patients on BAT (consisting of mostly ruxolitinib) are treated for life. In response to clarification, the company presented an analysis whereby responders (varying proportion) are assumed to remain on treatment, but assumed that non-responder stop treatment as per TTD. As described in Section 4.3.4.4, while the ERG considers that non-responders should also be continued on treatment, this analysis is presented for transparency and completeness. For this analysis, it is further assumed that 88.5% of patients continue fedratinib (this was done for fairness as 88.5% in the comparator arm are on ruxolitinib). It should be noted that this analysis is generated based on the updated company' model sent following clarification. The ERG was not able within the time available to check whether this analysis is implemented correctly.

- **Exploratory analysis 6: Assumption that 88.5% of fedratinib non-responders and responders continue treatment**

In this analysis, both responders and non-responders to fedratinib (88.5%) are assumed to be continue treatment for life, as per assumption for BAT. This is the ERG preferred assumption, given constraints imposed by the model structure and logic.

4.5 Conclusion of the cost-effectiveness section

The company's searches did not identify any economic analyses for fedratinib for the treatment of adult patients with MF previously treated with ruxolitinib.

The CS presents the methods and results of a *de novo* individual model to assess the cost-effectiveness of fedratinib versus BAT (comprising mostly of ruxolitinib) in patients with adult patients MF previously treated with ruxolitinib that are either resistant or intolerant to ruxolitinib. Incremental health gains, costs and cost-effectiveness are evaluated over a lifetime horizon from the perspective of the NHS and PSS, with health outcomes and costs discounted at a rate of 3.5%. The model includes a net price for fedratinib. The PAS for ruxolitinib is not considered.

The economic model is comprised of five key health states (i) on JAK, (ii) on BAT, (iii) AML transformation, (iv) palliative care and (v) death. Patients are further separated onto (i) early death, (ii) early discontinuer, (iii) non-responders at 24 weeks and (iv) responders at 24 weeks. Efficacy data (response rate, OS, adverse event frequency) for fedratinib is taken from JAKARTA-2 directly. OS for fedratinib and comparator is taken from two separate sources and is not adjusted for differences in patients characteristics, so the final comparison is a naïve indirect comparison. OS for the comparator

arm is taken from studies including people who discontinued ruxolitinib and were no longer treated with ruxolitinib. Health utility values for responders and non-responders were estimated using a mixed effect model fitted to MF-8D estimated in JAKARTA-2. Resource use were derived from a mix of sources in patient naïve to ruxolitinib. Scenario analysis are conducted assuming (a) a stopping rule at Week 24 and (b) OS estimated through surrogacy relationship based on response rate.

The deterministic version of the company's base case model suggest that fedratinib generate an additional [redacted] QALYs at an additional cost of [redacted] per patient compared with BAT (comprising mostly of ruxolitinib); the corresponding ICER is £13,905 per QALY gained. The probabilistic version of the model generate an additional [redacted] QALYs at an additional cost of [redacted] per patient compared with BAT (comprising mostly of ruxolitinib); the corresponding ICER is £10,384 per QALY gained.

The ERG's critical appraisal identified several issues relating to the company's model and the evidence used to inform its parameters. While some of these issues are debatable and may reflect matters of subjective opinion, others reflect more serious underlying problems regarding the conceptualisation and implementation of the model and the use of evidence to inform the model's parameters. Consequently, the ERG has serious doubts regarding the validity of the results presented within the CS and would advise considerable caution in their interpretation and use in informing decision-making. These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethinking of the model's logic. As such, the impact of these issues on the expected cost-effectiveness of fedratinib is not clear.

Key issues identified by the ERG included; (i) concerns with the modelling approach and health states, (ii) inconsistent assumptions between arms, (iii) approach to model OS and estimated survival differences in the absence of evidence to support this, (iv) the lack of face validity for the scenarios assuming a stopping rule or estimation of OS through surrogacy, (iv) questionable assumptions regarding cost for ruxolitinib, (v) lack of reliability of response rate and duration of response and (vi) questionable assumptions regarding HRQoL.

The ERG undertook a set of exploratory analyses, which taken together, comprise the ERG's preferred analysis. It should be noted that analyses were limited in nature by the company's modelling structure and approach and therefore the ERG express caution.

These included: (i) disabling options in the model (AML transformation, palliative care, DoR from week 24, wastage for ruxolitinib, utility estimate without effect of gender, thiamine supplementation), (ii) changes to parametric distributions (exponential distribution for TTD for non-responders and

generalised gamma distribution for OS for fedratinib), (iii) assumption of no difference in survival between treatment arms in the absence of robust evidence suggesting the contrary, (iv) adjustment to inputs so that model's prediction match original input (TTD for fedratinib) and (v) use of consistent assumption between the comparator and fedratinib arm (duration on treatment).

The ERG preferred base-case analysis suggest that the deterministic ICER is £2,959,869 per QALY gained when assuming that both responders and non-responders (88.5%) continue on treatment after TTD in the trial (more realistic scenario). The ICER was £444,999 per QALY gained when this assumption is not assumed. The ICER will be higher when the confidential PAS for ruxolitinib is taken into account.

Owing to the ERG's concerns regarding the robustness of the company's model, the results generated using the company's model, including the ERG's exploratory analyses, should be interpreted with caution. However, the exploratory analysis conducted by the ERG illustrate the likely impact on the ICER and that a complete rethink is required.

5 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

This section presents the results of the ERG's exploratory analyses. As previously described, the ERG has serious doubts regarding the validity of any results generated using the CS economic model[1, 37] and would advise considerable caution in their interpretation and use in informing decision-making. These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethink of the model's logic. As such, the impact of these issues on the expected cost-effectiveness of fedratinib is not entirely clear.

ICERs presented in this section are therefore subject to several limitations, with the analysis being limited by the model functionality/logic. Despite uncertainty, the ERG considers the ICERs presented in this section to be more realistic (given the model submitted) compared with those presented in the CS. Given that issues are intertwined with each other, the cumulative impact of each change is presented (Table 33). Results are presented in this section using ruxolitinib list price (a confidential PAS is available for ruxolitinib).

The first two amendments had little impact on the ICERs. When assuming no survival gain, the ICER for fedratinib improve, with fedratinib is dominant (e.g is more effective but associated with less costs). However, as previously described, there is an error in the model in that TTD predicted by the model does not match the original input. When TTD for fedratinib is corrected so that it matches the original input (TTD adjusted using a HR), the ICER increases to £444,999 per QALY gained.

When assessing the impact of using similar assumptions between arms regarding the duration on treatment (e.g treatment for life), the ICER increases to £1,382,748 per QALY gained when assuming only responders (88.5%) continue on treatment after the TTD in the trial and £2,959,869 per QALY gained when assuming that both responders and non-responders (88.5%) continue on treatment after TTD in the trial. This is considered a more realistic scenario (given the assumption made by the company for BAT [mostly ruxolitinib]) and therefore reflect the ERG's preferred base-case. Again, it should be noted that these analyses are limited by the model functionality and current model's logic. Although exploratory in nature, given limitations of the current CS economic model, these analyses shows that the ICER for fedratinib against BAT is very sensitive to, (1) the level of survival gain assumed (evidence lacking at present to support this assumption), (2) the correct modelling and estimation for TTD (inappropriate modelling approach) and (3) assumptions about how long patients remain on treatment with fedratinib.

Table 33: Exploratory ICERs generated by the ERG (cumulative impact)

	Cost		QALYs		Inc Cost	Inc QALYs	ICER
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Base-case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**£11,645**
Exploratory analysis 1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£8,477
Exploratory analysis 1-2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£8,303
Exploratory analysis 1-3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Dominant
Exploratory analysis 1-4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£444,999
Exploratory analysis 1-5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£1,382,748
Exploratory analysis 1-6 (ERG's preferred base-case)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**£2,959,869**

The ERG highlights that a number of additional issues could not be resolved which could have the effect of increasing ICER. For instance,

1) for patients on BAT after fedratinib (if treatment is not assumed to be continued for life), the ERG believes that HRQoL needs to be lower than the baseline as these patients would have failed 2 JAK and therefore they are likely to receive supportive care. The ERG does not consider it appropriate to assume utility value for BAT to be the same whether patients are treated or not with JAK.
2) The cost for ruxolitinib is highly uncertain and is a key driver. It is unclear if the platelet count distribution in JAKARTA-2 is representative of patients in UK practice. The ERG notes that costs for ruxolitinib would be lower if the platelet count distribution from SIMPLIFY-2 and other sources of evidence identified were used,
3) The model uses response assessed using spleen volume. A larger proportion of patients in the comparator arm would be classified as responders if spleen length (by palpitation) was used. The ICER would increase.
4) The impact of fedratinib on resource use is unknown. It is unclear to the ERG in what direction the ICER would change.
5) Finally, ICERs are also presented using ruxolitinib list price (ICER would increase if the confidential PAS for ruxolitinib was considered).

6 END OF LIFE

Two criteria needs to meet to satisfy NICE End of Life (EoL) criteria:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment.

The CS makes the case that patients have a short life expectancy referencing a number of studies in patients that have been treated with ruxolitinib. The CS states that these studies show a median OS of 13–16 months and these estimates are likely to be even lower in the intermediate-2 and high-risk population. The CS also comments on predictions from the economic model to justify the improvement in survival of more than 3 month. The CS states that the CS economic model predict that fedratinib provide an additional life year of 0.85 years compared with BAT. The CS further reference evidence from JAKARTA to support a survival gain.

While the ERG recognises the large unmet need for this population, it is the ERG's view that it is debatable whether the population entering the economic model (as defined by the company – patients on suboptimal ruxolitinib) meet EoL criteria for the following reasons:

- The studies quoted by the company are conducted in people who discontinued ruxolitinib. This is not in line with the population entering the economic model (89% of whom continued on ruxolitinib, at the point where they "should" have been discontinued);
- The EoL criteria is not met in the company's economic model (for its base-case) as it predicts a mean LY for BAT over 2 years ([redacted] years). It should be noted that this increases even further under the ERG preferred base-case;
- It is also unclear if there is, if any, and the magnitude of, survival gain that would be observed with fedratinib given the absence of head-to-head trial data against an appropriate comparator. The CS economic model generates a gain in survival in the base-case (however, evidence from JAKARTA-2 are naively compared with studies in the literature conducted in a different population). The CS quotes evidence in ruxolitinib naïve patient from JAKARTA to support the assumption of survival gain. The ERG noted that the HR for fedratinib 500mg vs. placebo in JAKARTA[14, 25] was [redacted]). As described in Section 4.3.4.6, it is the ERG's view that evidence from JAKARTA, SIMPLIFY-2 and COMFORT-trials do not support the assumption of a difference in survival.

7 OVERALL CONCLUSIONS

Based on one, single-arm study, the company's systematic review of clinical effectiveness suggests fedratinib to be significantly effective for treating splenomegaly and symptoms in myelofibrosis on spleen volume reduction ≥35% measured by MRI/CT at end of treatment cycle six (primary outcome) and cycle three (secondary outcome).

Based on the same single-arm study, the company's systematic review of clinical effectiveness also indicated that treatment with fedratinib is also associated with: patients achieving a duration of spleen response longer than nine months, a reduction in spleen volume, with an average reduction of one-third; reductions in spleen size by palpation; and an improvement in Total Symptom Score with more than a quarter achieving the clinically meaningful threshold for response of ≥ 50% reduction (secondary outcomes) (secondary outcomes). Based on the same single-arm study, the company's systematic review of clinical effectiveness indicates that overall survival data are immature. However, the company implies that fedratinib delivers a survival gain.

Quality-of-life (assessed using the EORTC-QLQ C30) and patient reported outcomes (assessed using MF-SAF) indicate that treatment with fedratinib is also associated with improvements in both HRQoL and symptoms of myelofibrosis and patients' symptom response rates; based on the same single-arm study.

All patients in the same single-arm study had at least one treatment-emergent adverse event (TEAE), with seven (7%) having a TEAE leading to death.

The ERG's critique of the clinical effectiveness evidence identified that study quality of the single-arm study was appropriately assessed by the company using an appropriate quality assessment instrument. However, although the single-arm study quality was adequate, the ERG notes the methodological limitations of single-arm study design for assessing clinical effectiveness. As such, the ERG considers that the results from this single-arm study should be interpreted with caution.

No pair-wise meta-analysis could be undertaken by the company, and the ERG has some concerns with the indirect treatment comparison (ITC) analyses that were performed by the company.

The company undertook an ITC to compare fedratinib with best available therapy (BAT - which is a comparator in the NICE Scope), using evidence from two randomised controlled trials of JAK2 inhibitors compared to BAT. Inclusion criteria in these studies appear different. The ERG's critique of the ITC methods identified concerns with the adjustment for prognostic factors and treatment effect

modifiers, differences in the populations across studies that cannot be adjusted for by matching of baseline characteristics. The ERG considers that, although the limitations of the ITC were described, that the extent of potential bias or uncertainty were not quantified. Furthermore, ERG has some concerns with some aspects of the methods used to identify prognostic factors and treatment effect modifiers. Whilst adjustments were made for some prognostic factors, the ERG has concerns with one of the RCTs used in the ITC, in which there was a much greater proportion of patients who were transfusion dependent, and has concerns if the population is representative of the target population. Other concerns of the ERG include the estimates of relative treatment effect (presented on the absolute risk scale rather than the odds ratio, or log-odds ratio scale) and that the risk difference is unlikely to be generalisable and that the primary purpose of back-transforming to treatment-specific absolute risks is for use in the economic model. Consequently, the ERG suggests that the way the ITC results have been used is inappropriate. The ERG also notes that an effect on overall survival is also not proven.

The ERG's critical appraisal identified several important issues relating to the company's model and the evidence used to inform its parameters. While some of these issues are debatable and may reflect matters of subjective opinion, others reflect more serious underlying problems regarding the conceptualisation and implementation of the model and the use of evidence to inform the model's parameters. Consequently, the ERG has serious doubts regarding the validity of the results presented within the CS and would advise considerable caution in their interpretation and use in informing decision-making. These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethinking of the model's logic. As such, the impact of these issues on the expected cost-effectiveness of fedratinib is not clear.

The ERG's preferred deterministic ICER, given the current modelling structure and approach is £2,959,869 per QALY gained when assuming that both responders and non-responders (88.5%) continue on treatment after TTD in the trial; this is significantly higher compared with the CS estimate of £13,605 per QALY gained. The deterministic ICER is reduced to £444,999 per QALY gained, if patients are assumed to discontinue as per TTD in the trial. However, the ERG re-iterate that is inconsistent with the key assumption for the comparator and consider this to less realistic. The ERG further notes that different assumptions be used between treatment, it is not appropriate to assume that patients who discontinued after 2 JAKs would have the same benefit in HRQoL as patients currently on JAK.

The ERG was not able to account for all uncertainties in the economic model due to the constraint imposed by the model structure and logic, and therefore the ICERs could be higher.

Owing to the ERG's concerns regarding the robustness of the company's model, the results generated using the company's model, including the ERG's exploratory analyses, should be interpreted with caution. However, the exploratory analysis conducted by the ERG illustrate the likely impact on the ICER and the uncertainty around the analysis and its results.

8 REFERENCES

1. Celgene. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; 2020.
2. National Institute for Health and Care Excellence. Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis - Final Scope. May 2020. https://www.nice.org.uk/guidance/gid-ta10335/documents/final-scope (Accessed 25 January 2019).
3. Reilly JT, McMullin MF, Beer PA, Butt N, Conneally E, Duncombe A*, et al.* Guideline for the diagnosis and management of myelofibrosis. *Br J Haematol* 2012;158:453-71.
4. Romano M, Sollazzo D, Trabanelli S, Barone M, Polverelli N, Perricone M*, et al.* Mutations in JAK2 and Calreticulin genes are associated with specific alterations of the immune system in myelofibrosis. *Oncoimmunology* 2017;6:e1345402.
5. Song M-K, Park B-B, Uhm J-E. Understanding Splenomegaly in Myelofibrosis: Association with Molecular Pathogenesis. *International journal of molecular sciences* 2018;19:898.
6. Moulard O, Mehta J, Olivares R, Iqbal U, Mesa RA. Epidemiology of Myelofibrosis (MF), Polycythemia Vera (PV) and Essential Thrombocythemia (ET) in the European Union (EU). *Blood* 2012;120:1744-.
7. Myelofibrosis. MV. No date.
8. Michaelis LC. Risk stratification in Myelofibrosis: The quest for simplication. *Haematologica* 2017;102:2-3.
9. Mascarenhas J, Hoffman R, Talpaz M, Gerds AT, Stein B, Gupta V*, et al.* Pacritinib vs Best Available Therapy, Including Ruxolitinib, in Patients With Myelofibrosis: A Randomized Clinical Trial. *JAMA Oncol* 2018;4:652-9.
10. Harrison CN, Vannucchi AM, Platzbecker U, Cervantes F, Gupta V, Lavie D*, et al.* Momelotinib versus best available therapy in patients with myelofibrosis previously treated with ruxolitinib (SIMPLIFY 2): a randomised, open-label, phase 3 trial. *Lancet Haematol* 2018;5:e73-e81.
11. Harrison CN, Schaap N, Vannucchi AM, Kiladjian J-J, Tiu RV, Zachee P*, et al.* Janus kinase-2 inhibitor fedratinib in patients with myelofibrosis previously treated with ruxolitinib (JAKARTA-2): a single-arm, open-label, non-randomised, phase 2, multicentre study. *Lancet Haematol* 2017;4:e317-e24.
12. DeAngelo D. Diagnosis/risk classification for primary myelofibrosis. 2017. https://www.targetedonc.com/case-based-peer-perspectives/myeloproliferative-neoplasms/deangelo-highrisk-mf/diagnosis-risk-classification-for-primary-myelofibrosis (Accessed 5 December 2019).
13. National Institute for Health and Care Excellence. Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. Technology appraisal guidance [TA386]. 2016. https://www.nice.org.uk/guidance/TA386/chapter/1-Recommendations (Accessed 13 November 2019).
14. Pardanani A, Harrison C, Cortes JE, Cervantes F, Mesa RA, Milligan D*, et al.* Safety and efficacy of fedratinib in patients with primary or secondary myelofibrosis: a randomized clinical trial. *JAMA Oncol* 2015;1:643-51.
15. Celgene. Summary of Product Characteristics (SPC). Inrebic® (fedratinib) 100mg hard capsules. 2020. TBD (Accessed August 2020).
16. NCT03755518. A Safety Trial of Fedratinib in Subjects With DIPSS, Intermediate or High-Risk Primary Myelofibrosis, Post-Polycythemia Vera Myelofibrosis, or Post-Essential Thrombocythemia Myelofibrosis and Previously Treated With Ruxolitinib With Concomitant Luspatercept for Subjects With Anemia (FREEDOM). *Journal* 2020. https://clinicaltrials.gov/ct2/show/NCT03755518 [Accessed 8 August 2020]
17. NCT03952039. An Efficacy and Safety Study of Fedratinib Compared to Best Available Therapy in Subjects With DIPSS-intermediate or High-risk Primary Myelofibrosis, Post-polycythemia Vera Myelofibrosis, or Post-essential Thrombocythemia Myelofibrosis and Previously Treated With Ruxolitinib (FREEDOM2). *Journal* 2020. https://clinicaltrials.gov/ct2/show/NCT03952039 [Accessed 4 August 2020]

18. Celgene. Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis [ID1501]. Clarification response; 2020.
19. Celgene. ARD12181 (JAKARTA2): A Phase II, multicenter, open-label, single-arm study of SAR302503 in subjects previously treated with ruxolitinib and with a current diagnosis of intermediate or high-risk primary myelofibrosis, post-polycythemia vera myelofibrosis, or post-essential thrombocythemia myelofibrosis (clinical study report). In; 2018.
20. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-84.
21. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A*, et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009;114:937-51.
22. Cervantes F, Dupriez B, Pereira A, Passamonti F, Reilly JT, Morra E*, et al.* New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood* 2009;113:2895-901.
23. Harrison C, Schaap N, Vannucchi A, Kiladjian JJ, Jourdan E, Silver R*, et al.* Fedratinib in patients with myeloproliferative neoplasm-associated myelofibrosis previously treated with ruxolitinib: a reanalysis of the Phase 2 JAKARTA-2 study. In: European Hematology Association. Amsterdam: Netherlands; 2019.
24. Harrison CN, Schaap N, Vannucchi AM, Kiladjian JJ, Jourdan E, Silver RT*, et al.* Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the JAKARTA2 study using stringent criteria for ruxolitinib failure. *American Journal of Hematology* 2020.
25. Celgene. Overall survival estimates from JAKARTA and JAKARTA-2. In; 2019.
26. Celgene. JAKARTA-2 patient level data analysis. In; 2020.
27. Celgene. Myelofibrosis Advisory Board Meeting Report. In; 2020.
28. Celgene. ARD12181 (JAKARTA2): A Phase II, multicenter, open-label, single-arm study of SAR302503 in subjects previously treated with ruxolitinib and with a current diagnosis of intermediate or high-risk primary myelofibrosis, post-polycythemia vera myelofibrosis, or post-essential thrombocythemia myelofibrosis (clinical study report). In; 2018.
29. Celgene. Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis [ID1501]. Clarification response No. 2; 2020.
30. Celgene-2019 data on file [AIC]. Targeted Literature Review: An Exploration of Surrogacy between Treatment Response and Overall Survival in Myelofibrosis. In; 2019.
31. Kirby S, Burke J, Chuang-Stein C, Sin C. Discounting phase 2 results when planning phase 3 clinical trials. *Pharmaceutical Statistics* 2012;11:373-85.
32. Consilient Health. Pentosan polysulfate sodium for treating bladder pain syndrome [ID1364]; 2019.
33. Celgene. External clinical validation exercises - fedratinib; 2019.
34. National Institute for Health and Care Excellence. Single Technology Appraisal (STA): User guide for company evidence submission template; 2017.
35. Efthimiou O, Mavridis D, Nikolakopoulou A, Rucker G, Trelle S, Egger M*, et al.* A model for meta-analysis of correlated binary outcomes: The case of split-body interventions. *Stat Methods Med Res* 2019;28:1998-2014.
36. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B, Collaborators. Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol* 2018;98:133-43.
37. Celgene. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501 - Economic model. 2020.
38. Schain F, Vago E, Song C, He J, Liwing J, Lofgren C*, et al.* Survival outcomes in myelofibrosis patients treated with ruxolitinib: A population-based cohort study in Sweden and Norway. *Eur J Haematol* 2019;103:614-9.
39. Kuykendall AT, Shah S, Talati C, Al Ali N, Sweet K, Padron E*, et al.* Between a rux and a hard place: evaluating salvage treatment and outcomes in myelofibrosis after ruxolitinib discontinuation. *Ann Hematol* 2018;97:435-41.

40. Palandri F, Breccia M, Bonifacio M, Polverelli N, Elli EM, Benevolo G, *et al.* Life after ruxolitinib: Reasons for discontinuation, impact of disease phase, and outcomes in 218 patients with myelofibrosis. *Cancer* 2019; 10.1002/cncr.32664.
41. Tefferi A, Cervantes F, Mesa R, Passamonti F, Verstovsek S, Vannucchi AM, *et al.* Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. *Blood* 2013;122:1395-8.
42. Mesa RA, Li CY, Ketterling RP, Schroeder GS, Knudson RA, Tefferi A. Leukemic transformation in myelofibrosis with myeloid metaplasia: a single-institution experience with 91 cases. *Blood* 2005;105:973-7.
43. Tefferi A, Lasho TL, Jimma T, Finke CM, Gangat N, Vaidya R, *et al.* One thousand patients with primary myelofibrosis: the mayo clinic experience. In: Mayo Clinic Proceedings: Elsevier; 2012: 25-33.
44. Rafia R, Mukuria CW. Exploratory Psychometric Analysis of The Eq-5d In A Myelofibrosis Population. *Value in Health* 2015;18:A211.
45. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, *et al.* Deriving a Preference-Based Measure for Cancer Using the EORTC QLQ-C30. *Value in Health* 2011;14:721-31.
46. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health* 2015;18:846-55.
47. Mukuria CW, Brazier J, Rafia R. Does the generic cancer outcome measure eortc qlq-c30 work in myelofibrosis? *Value in Health* 2015;18:A210-A1.
48. Pan F, Peng S, Fleurence R, Linnehan JE, Knopf K, Kim E. Economic Analysis of Decitabine Versus Best Supportive Care in the Treatment of Intermediate- and High-Risk Myelodysplastic Syndromes From a US Payer Perspective. *Clinical Therapeutics* 2010;32:2444-56.
49. Ara R, Brazier JE. Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. *Value in Health* 2010;13:509-18.
50. Mead AJ, Milojkovic D, Knapper S, Garg M, Chacko J, Farquharson M, *et al.* Response to ruxolitinib in patients with intermediate-1-, intermediate-2-, and high-risk myelofibrosis: results of the UK ROBUST Trial. *Br J Haematol* 2015;170:29-39.
51. Haematological Malignancy Research N. Haematological Malignancy Research Network: Myelofibrosis Audit. 2012. https://www.hmrn.org/Download.aspx?target=document&id=3 (Accessed 6 March 2020).
52. Haematological Malignancy Research N. Haematological Malignancy Research Network: Clinical Management, Resource Utilisation and Outcome in Primary and Secondary Myelofibrosis. In; 2020.
53. Al-Ali HK, Griesshammer M, le Coutre P, Waller CF, Liberati AM, Schafhausen P, *et al.* Safety and efficacy of ruxolitinib in an open-label, multicenter, single-arm phase 3b expanded-access study in patients with myelofibrosis: a snapshot of 1144 patients in the JUMP trial. *Haematologica* 2016;101:1065-73.
54. National Health Service. Provider to provider services. 2019. https://www.uclh.nhs.uk/aboutus/wwd/Documents/Provider%20to%20Provider%20Tariff%202018-19.pdf (Accessed 27 May 2020).
55. Wang HI, Aas E, Howell D, Roman E, Patmore R, Jack A, *et al.* Long-term medical costs and life expectancy of acute myeloid leukemia: a probabilistic decision model. *Value Health* 2014;17:205-14.
56. Round J, Jones L, Morris S. Estimating the cost of caring for people with cancer at the end of life: A modelling study. *Palliat Med* 2015;29:899-907.
57. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London, UK; 2013.
58. EMC. Summary of Product Characteristics Updated 14-Aug-2020. Ruxolitanib (Jakavi 10mg Tablets) Novartis Pharmaceuticals UK Ltd.; 2020. TBD (Accessed TBD).
59. Mehra M, Potluri R, He J, Wang G, Mundle S, Bussolari J. Characterization of disease, treatment patterns, and outcomes of patients with myelofibrosis: analysis of two United States

commercial claims databases. In: American Society of Hematology (ASH) 58th Annual Meeting. San Diego, California; 2016.

60. Oncology Sierra. Targeted Hematology and Oncology Therapeutics. 2018. http://investor.sierraoncology.com/download/Sierra+Analyst+%26+Investor+Call+Presentation_Final.pdf (Accessed 22 June 2020).
61. Palandri F, Palumbo GA, Bonifacio M, Tiribelli M, Benevolo G, Martino B*, et al.* Baseline factors associated with response to ruxolitinib: an independent study on 408 patients with myelofibrosis. *Oncotarget* 2017;8:79073-86.
62. Vannucchi AM, Kantarjian HM, Kiladjian JJ, Gotlib J, Cervantes F, Mesa RA*, et al.* A pooled analysis of overall survival in COMFORT-I and COMFORT-II, 2 randomized phase III trials of ruxolitinib for the treatment of myelofibrosis. *Haematologica* 2015;100:1139-45.
63. Verstovsek S, Kantarjian HM, Estrov Z, Cortes JE, Thomas DA, Kadia T*, et al.* Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. *Blood* 2012;120:1202-9.
64. Reilly JT, McMullin MF, Beer PA, Butt N, Conneally E, Duncombe AS*, et al.* Use of JAK inhibitors in the management of myelofibrosis: a revision of the British Committee for Standards in Haematology Guidelines for Investigation and Management of Myelofibrosis 2012. *British Journal of Haematology* 2014;167:418-20.
65. Animesh Pardanani RS, Moshe Talpaz, Jorge Cortes, Jason Gotlib, Catriona Jamieson, Shelonitda Rose, Tymara Berry, Jun Zhang, Ayalew Tefferi. Long-Term Safety of Fedratinib in Patients with Intermediate- or High-Risk Myelofibrosis (MF). *EHA25; Conference abstract EP1096 (25th Virtual conference)* 2020.

9 APPENDICES

Appendix 1. Example configurations for SVR responders and TSS responders in the ITC

Appendix 1 Table 1

	SVR			Total
TSS		Y	N	
	Y	3	0	3
	N	2	25	27
Total		5	25	30

Appendix 1 Table 2

	SVR			Total
TSS		Y	N	
	Y	2	1	3
	N	3	24	27
Total		5	25	30

Appendix 1 Table 3

	SVR			Total
TSS		Y	N	
	Y	1	2	3
	N	4	23	27
Total		5	25	30

Appendix 1 Table 4

	SVR			Total
TSS		Y	N	
	Y	0	3	3
	N	5	22	27
Total		5	25	30

Fedratinib for splenomegaly and symptoms in myelofibrosis. A Single Technology Appraisal

ERG commentary on the company's addendum submission to NICE

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK Rachid Rafia, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK
Correspondence Author	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
Date completed	09/04/2021

2. Introduction

This document has been produced in response to the Celgene addendum submission to NICE,[1] and includes analyses that align with those (exploratory) requested by the Committee for Medicinal Products for Human Use (CHMP),[2] in that that patients should be counted as non-responders if they achieved a ≥ 35% reduction in spleen volume from baseline to the End of Cycle 6 (spleen response – primary endpoint) with a dose higher than 400 mg daily of fedratinib. The company's addendum submission to NICE[1] also included an update to the net price of fedratinib.

This document also sets out the ERG's commentary on the company's addendum to NICE submission (following the CHMP); which included a submission of updated evidence for response rates and a revised economic model.

3. Additional evidence submitted by the company and summary of model's update

- **Additional evidence: Updated response rates for fedratinib based on CHMP analysis**

The CHMP requested an exploratory analysis where patients who responded after their dose were up-titrated within the first 6 cycles to be counted as non-responders. Response rates for the both the ITT population (including intermediate-1) and the intermediate-2/high risk population, using the original and updated response definition (CHMP analysis) are summarised in Table 1.

Table 1: Response rates using original and updated definition (CHMP analysis)

Symptom response **End of cycle 6**	**ITT population - Original (N=97)**	**ITT population – CHMP update (N=97)**	**Intermediate-2/ high-risk disease – Original (N=81)**	**Intermediate-2/ high-risk disease – CHMP update (N=81)**
SVR n (%)	[redacted]	[redacted]	[redacted]	[redacted]
Symptom response n (%)	[redacted]	[redacted]	[redacted]	[redacted]
Symptom or SVR response n (%)	[redacted]	[redacted]	[redacted]	[redacted]
Key: ITT, intention-to-treat; LOCF, last observation carried forward				

- **Summary of model's update since technical engagement**

The economic model has been subsequently updated by the company to include the updated response definition aligning with the exploratory CHMP request. It should be noted that the addendum[1] was submitted by the company after the ERG's technical engagement response document.[3] Consequently, the addendum submitted by the company includes additional updates to the model, made in response to the ERG's technical response document (including correction of errors identified).

The following key model input updates are described by the company following technical engagement:

1. Change to the response rates for fedratinib and update to response ITC based on exploratory CHMP analysis,
2. Estimate of utility values based on response definition requested by CHMP,
3. Update to time to treatment discontinuation and overall survival separated by response status,
4. Inclusion of dose intensities for fedratinib,
5. Update to net price of fedratinib

4. ERG's commentary on company's addendum to NICE

a. Updated response definition and ITC

The ERG understands that the EMA[2] asked the company to provide an exploratory analysis re-classifying patients who were up-titrated (>400mg) as non-responders. However, the ERG notes that there is no explicit mention in the EMA marketing authorisation that patients treated with fedratinib cannot receive more than 400 mg. The ERG further notes that the EMA's additional analysis appears to be exploratory.

Consequently, the ERG requested clarification from the company for the exact wording of the marketing authorisation that up-titration (dose > 400 mg) with fedratinib is not allowed and a copy of the protocol for FREEDOM-2.[4] In its response (see addendum clarification question C1[4]), the company states that the recommended dose of fedratinib is 400 mg daily, the SmPC does not mention dose escalation for patients with an insufficient spleen and/or symptom response at the 400 mg dose (but was explicitly mentioned in JAKARTA-2), the CHMP report states that it was not clear if up-titration with fedratinib provided additional clinical benefit, and that the dose in FREEDOM-2 (which included 5 UK centres) cannot exceed 400 mg daily.

While the ERG considers the response from the company to be generally satisfactory, the ERG notes that the possibility that patients would not be up-titrated in practice cannot be ruled out. The ERG further notes that in the FREEDOM-2 protocol, it is explicitly stated that the fedratinib dose cannot exceed 400 mg daily. However, it remains unclear if this reflects how fedratinib will be used in UK practice.

In addition to updates to the response definition for fedratinib, the company also updated the ITC using the same methods as previously described.[3, 5] Limitations with the ITC have been previously described in the original ERG report,[5] and ERG's technical engagement response document.[3]

b. General model structure, extrapolation and assumption of survival gain

The ERG re-iterates that the model is overly complicated due to choices made by the company to separate patients using response rates, to use an individual-based approach and the evidence used. All of these decisions add to the challenges associated with the single arm nature of the JAKARTA-2 trial, along with its small sample size, the impact of clinical hold, and censoring patients who got up-titrated in the updated analysis.

In the company's updated base-case, Overall survival (OS) for fedratinib and BAT are modelled independently from each other, with OS for fedratinib estimated based on (i) response rate, (ii) time to treatment discontinuation separated by response status and (iii) time to death following treatment discontinuation also separated by response status. In contrast, OS for BAT is estimated by the company by fitting a parametric distribution to Schain et al (2019) in patients who discontinued ruxolitinib (e.g not a population that is continued on suboptimal ruxolitinib as per the population entering the economic model [there is therefore a disconnect between costs and effectiveness]). Consequently, the difference in OS predicted in the company's model are not attributable to the differences in response rates, but to the different approaches used and inappropriate use of evidence for OS for BAT (88.5% continued on ruxolitinib) in patients who discontinued ruxolitinib.

Different approaches to treatment discontinuation are used. Patients initiated on fedratinib are assumed by the company to discontinue fedratinib as per the discontinuation rate in JAKARTA-2 and receive non-JAK treatments (mostly HU – low cost) until moving to supportive care.
In contrast, patients initiated in the BAT arm (composed of 85.5% ruxolitinib) are assumed remain on the same treatments (composed of mostly ruxolitinib – high cost) until supportive care.

An in-depth description of the ERG's critique of the company structure/assumption is available in the original ERG report,[5] and technical engagement response documents[3].

- **Assumption of survival difference and treatment received post-fedratinib discontinuation**

At technical engagement,[6] the ERG requested that the company provide a matched indirect comparison for OS between SIMPLIFY-2 and JAKARTA-2, to adjust for potential differences in baseline

characteristics. Results were provided by the company and have been replicated below in Table 2. The analysis indicates that after matching, OS up to 24 weeks in patients initiated on fedratinib in JAKARTA-2 is not better than OS in a similar population treated with BAT in SIMPLIFY-2, and could in fact be worse when adjusting for important variables (such as platelet count and transfusion dependence). Further details are available in the ERG's technical engagement document.[3]

Table 2: Exploratory OS MAICs with SIMPLIFY-2 BAT arm in first 24 weeks (reproduction of Table 3 in TE company's response)

Method	HR (95% CI)	JAKARTA-2 N / ESS
Naïve	[redacted]	[redacted]
MAIC (matching on DIPSS)	[redacted]	[redacted]
MAIC (DIPSS: matching on age and haemoglobin)	[redacted]	[redacted]
MAIC (DIPSS plus: matching on age, haemoglobin, transfusion dependence and platelet counts)	[redacted]	[redacted]

It is evident from Table 2 that the survival difference predicted by the company's model does not align with the evidence from SIMPLIFY-2 (that is used by the company to justify a difference in response rates).

The company further assumes that patients initiated on BAT (88.5% ruxolitinib) remain on treatment for life (until reaching supportive care) incurring high costs, while patients initiated fedratinib discontinue using the discontinuation rate observed in JAKARTA-2 and move onto non-JAK BAT (assumed to be 66.7% hydroxyurea, 16.7% prednisolone, 16.7% prednisone) incurring low costs for the reminder of their life.

The clinical expert interrogated as part of the technical engagement[7] believed that following relapse, patients treated with fedratinib would either remain on fedratinib, switch to ruxolitinib or move to a clinical trial. This view was also shared by the ERG clinical advisors.[5] The ERG further believes that assuming patients on fedratinib remain on fedratinib (or ruxolitinib) is also consistent with the company's own argument for ruxolitinib (88.5% of treatment in the BAT arm), that patients are continued on ruxolitinib in the absence of alternative therapeutic options.

Consequently, in line of evidence available on survival, and clinical experts' statements, the ERG requested the company to provide two analyses that are deemed to be more realistic. These scenarios form the basis of the ERG's preferred base-case:

- **Analysis 1**: Assumes equal OS and time on treatment between fedratinib and BAT. Patients continue to receive fedratinib in post-fedratinib BAT using the same proportion of patients on ruxolitinib as the BAT arm.
- **Analysis 2**: Assumes equal OS and time on treatment between fedratinib and BAT. Patients receive ruxolitinib in post-fedratinib BAT, using the same proportion of patients on ruxolitinib as the BAT arm.

These two analyses were provided by the company as part of their addendum to NICE submission. [1, 4] The company notes that assuming patients receive ruxolitinib after fedratinib is outside its license and not recommended in the UK.

- **Uncertainty associated with long-term extrapolation**

The ERG notes that the company's decision to separate patients onto responders and non-responders, coupled with the small sample size (n=81), and the impact of clinical hold and censoring in patients who got up-titrated; increases the uncertainty when extrapolating survival curves. OS for fedratinib is a function of response rate, time to treatment discontinuation split by response status and time to death following discontinuation separated by response status.

- TTD extrapolation

The company selected the exponential and lognormal distribution for TTD for non-responders and responders respectively. The ERG notes that the choice of curves is very uncertain due to the short trial duration, and is confounded by the clinical hold and censoring for up-titration.

The ERG notes that the choice between parametric distributions is challenging and that other parametric distributions could be considered equally plausible. While the Gompertz distribution provided the most conservative estimate for TTD, this could be deemed to be as plausible as those selected by the company.

For transparency and completeness, and as this is uncertain, a scenario analysis is presented by the ERG assuming the Gompertz distribution for TTD for both responders and non-responders.

- Death following discontinuation

The company estimated the time to death following discontinuation separated by response status, increasing the uncertainty further. The ERG notes that the KM for time to death following discontinuation for responders is based on [redacted] patients only, where [redacted] event is observed (Figure 7 addendum to NICE submission[1]). It is unclear to the ERG whether outcomes for responders and non-responders following discontinuation would be different, and therefore uncertain.

5. Utility values

- **Estimate of utility values based on updated response definition**

The ERG has identified some inconsistencies in the estimation of utility values between the updated response definition and its original response definition. The ERG requested that the company clarify why the increment in utility values for non-responders is decreased ([redacted]) compared with the original response definition ([redacted]), when responders (who got up-titrated) are included in the non-responder group (when gender is not included in the utility model). The company (see addendum clarification question B2[4]) stated that "*for the addendum, utility analyses were updated with the new definition of response (at end of cycle 6) but were also further updated to include all available post-baseline utility values (rather than only those at end of cycle 3 and end of cycle 6). Therefore, the updated analyses are not directly comparable to the original analyses."*

Due to the lack of information, the ERG is not able to comment on this issue but notes that it is unclear why different approaches are used.

- **Assumption of no increment in utility values for non-responders on BAT**

Compared with its original submission, the company assumes no increment in utility values for non-responders on BAT in both its technical engagement response[6] and addendum.[1] The company states in its technical engagement response: "*since the utility data in the model is derived from JAKARTA-2 which was a single-arm study, the utility increment for 'non-responders' is reflective of a group of patients who did not meet the dichotomous responder criteria, but still achieved some clinical benefit. In contrast, non-responding patients receiving BAT in SIMPLIFY-2 often had increases in spleen size. Therefore, it was felt plausible that non-responders to BAT (and patients receiving BAT after fedratinib) should have no increment in utility*."

In its response to clarification (see addendum clarification question B2[4]), the company further states that "*the majority of non-responders in SIMPLIFY-2 as indicated by their TSS change from their*

baselines had a worsening of their symptoms…and therefore the 0-utility applied for non-responders to BAT would be reasonable".

The ERG notes that this is an area of uncertainty. The ERG is generally satisfied with the approach taken by the company and considers it plausible for no utility value increment for non-responders initiated on BAT (compared with non-responders on fedratinib) when looking at the waterfall plot of TSS in SIMPLIFY-2. However, the ERG re-iterates that this is an area of considerable uncertainty. The ERG re-iterates that there is important difference between the population included in JAKARTA-2 and SIMPLIFY-2. The ERG further notes that the MPN-SAF is used in SIMPLIFY-2 and is therefore less specific to MF. The washout period was also short in SIMPLIFY-2 (compared with JAKARTA-2); all of which may have contributed to the reduce benefit in symptoms observed in SIMPLIFY-2.

d. Inclusion of dose intensity for fedratinib and wastage for ruxolitinib

- **Inclusion of dose intensity in patients initiated on fedratinib**

Compared with its original submission to NICE,[8] the company includes in its addendum[1] dose intensity for fedratinib ([redacted]) to adjust for dose up-titration in JAKARTA-2. While this is uncertain, the ERG is generally satisfied with the company's approach (given the data available) if it is assumed patients cannot receive more than 400mg daily.

For the ERG's preferred analyses, where patients on fedratinib remain on treatment (as per assumption used for the BAT arm [88.5% ruxolitinib]), the company assumes a different dose intensity ([redacted]) in patients that are continued on 'suboptimal' fedratinib (e.g following fedratinib discontinuation from JAKARTA-2). This value is calculated by the company from the dose intensity in the subset of patients who did not get titrated (e.g patients who got up-titrated are removed altogether). The ERG notes that removing patients who got up-titrated (rather than capping the maximum dose to 400mg) is arbitrary and not appropriate. This is also not in line with the company's argument that the daily dose of fedratinib is 400 mg daily, and that that dose modifications are only referred in the SmPC in the context of managing treatment-emergent adverse reactions (see addendum clarification question C1[4]). The ERG believes that a dose intensity of [redacted]% is more plausible for patients that are maintained on fedratinib.

- **Inclusion of wastage for ruxolitinib**

Despite the ERG's previous comments,[5] the company's base-case includes 5% wastage for ruxolitinib. As highlighted in the ERG report,[5] this is not appropriate and not in line with the committee's conclusion in TA386.

e. Assumption that a proportion of patients receive no BAT treatment costs after fedratinib

For the scenarios requested by ERG where patients are assumed to remain on fedratinib or move to ruxolitinib, the company assumed that 12.5% of patients in the fedratinib arm not continued on JAK receive no other treatments and therefore accrue no drug cost. The ERG notes that this does not align with the assumption in the BAT arm.

f. Assumption of [redacted] AML transformation in patients initiated on fedratinib

In its original submission to NICE,[8] the company includes AML as a separate health state, and assumed the same rate of AML transformation between arms. This was justified by the company, stating that it is unclear if treatment influences the rate of progression to AML (CS, Table 29, page 93). This was considered reasonable by the ERG.[5] Following a series of programming errors by the company, AML is now included as an adverse event in the model as suggested by the ERG.[5]

However, the ERG notes that in the company's models sent at technical engagement[6] and in the addendum to NICE submission[1], arbitrary changes to the rate of AML (after the ERG report) have not been documented/described by the company and are favourable to fedratinib.

In its addendum submission to NICE,[1] the company assumes that [redacted] patients on fedratinib would transform to AML when restricting to the intermediate-2/high risk population in JAKARTA-2 (n=81; [redacted] event). However, in the company's original submission,[8] the rate of AML from the ITT population was used (n=97; [redacted] event). Also, the duration of follow-up in JAKARTA-2 is short ([redacted] weeks).

In contrast for BAT, the company takes the rate of AML from COMFORT-II (n=73; 4 events). In COMFORT-II, the duration of follow-up was considerably longer (101.0 weeks). No AML was reported in SIMPLIFY-2.

The ERG believes that the same rate of AML transformation should be used as it is unclear if treatment influences the rate of progression to AML, as argued by the company in its original submission to NICE.[8] Furthermore, as highlighted in the ERG report,[5] using the transformation rate from the long-term COMFORT trials for ruxolitinib (as per the company' notes in the economic model) for each arm is more appropriate as this is based on a longer follow-up (126.3 weeks) and number of patients (n=146; 5 events).

6. ERG's preferred base-case

The ERG's preferred base case is based on the analyses requested to the company assuming (1) no survival difference (in line with findings from the MAIC) and (2) that patients on fedratinib remain on either fedratinib or move to off-license ruxolitinib (e.g same assumption as patients initiated on BAT).

Compared with the company's analyses, the ERG's preferred base-case also includes the following amendments:

1. Inclusion of BAT drug costs for 12.5% not continued on JAKs,
2. Exclusion of gender from the regression model when estimating utility values,
3. Exclusion of wastage for ruxolitinib as per committee's conclusion in TA386,
4. Assumption of a dose intensity of [redacted]% for patients continued on "suboptimal" fedratinib,
5. Assumption of the same rate of AML transformation between arms based on the long-term COMFORT-trials,
6. Use AE rates from the ITT population (rather than restricted to the intermediate-2/high risk population)

The impact on the ICER associated with each individual change is presented, as well as the combined impact, which forms the ERG's preferred base-cases. It should be noted that the ICERs differ slightly from those reported by the company in its addendum submission to NICE, due to the correction of errors by the company following clarification from the ERG.[4]

The ERG re-iterates that the base-case assumptions in the company's model are not appropriate (e.g assumption of survival difference and inconsistent assumption between treatment arms around discontinuation). However, for transparency and completeness, the impact of changes in the company base-case are also reported below. Results should be interpreted with considerable caution as different approaches to discontinuation is used between arms.

- **ERG's preferred base-case 1: Assumption of no survival difference and 88.5% patients remain on fedratinib until supportive care (in line with assumption for BAT arm [88.5% ruxolitinib])**

	BAT			Fedratinib			Incremental			
	LY	QALY	Cost	LY	QALY	Cost	LY	QALY	Cost	ICER
Company analysis	**2.912**	[redacted]	[redacted]	**2.912**	[redacted]	[redacted]	**0.000**	[redacted]	[redacted]	**£61,513**
Analysis 1: Inclusion non-JAK costs	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£61,589
Analysis 2: Exclusion Gender from utility regression	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£63,187
Analysis 3: No wastage for ruxolitinib	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£81,997
Analysis 4: Dose intensity after FED	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£66,833
Analysis 5: Same AML rate (LT COMFORT-II)	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£83,247
Analysis 6: AEs rate from ITT population	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£61,767
ERG preferred base-case (analysis 1-6)	**2.912**	[redacted]	[redacted]	**2.912**	[redacted]	[redacted]	**0.000**	[redacted]	[redacted]	**£114,005**
Scenario analysis: ERG base-case + Gompertz for TTD	2.548	[redacted]	[redacted]	2.548	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£127,846

- **ERG's prefered base-case 2: Assumption of no survival difference and 88.5% patients move to ruxolitinib until supportive care (in line with assumption for BAT arm [88.5% ruxolitinib])**

	BAT			Fedratinib			Incremental			
	LY	QALY	Cost	LY	QALY	Cost	LY	QALY	Cost	ICER
Company analysis	**2.912**	[redacted]	[redacted]	**2.912**	[redacted]	[redacted]	**0.000**	[redacted]	[redacted]	**£67,173**
Analysis 1: Inclusion non-JAK costs	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£67,249
Analysis 2: Exclusion Gender from utility regression	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£69,001
Analysis 3: No wastage for ruxolitinib	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£82,987
Analysis 4: Dose intensity after FED	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£67,173
Analysis 5: Same AML rate (LT COMFORT-II)	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£89,189
Analysis 6: AEs rate from ITT population	2.912	[redacted]	[redacted]	2.912	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£67,443
ERG preferred base-case (analysis 1-6)	**2.912**	[redacted]	[redacted]	**2.912**	[redacted]	[redacted]	**0.000**	[redacted]	[redacted]	**£109,316**
Scenario analysis: ERG base-case + Gompertz for TTD	2.548	[redacted]	[redacted]	2.548	[redacted]	[redacted]	0.000	[redacted]	[redacted]	£121,392

- **Company's base-case: Survival difference and assumption that patients discontinue fedratinib early**

	BAT			Fedratinib			Incremental			
	LY	QALY	Cost	LY	QALY	Cost	LY	QALY	Cost	ICER
Company analysis	**2.394**	**1.359**	[redacted]	**2.912**	**1.839**	[redacted]	**0.518**	**0.480**	**£11,866**	**£24,736**
Analysis 1: Remove Gender utility regression	2.394	1.356	[redacted]	2.912	1.830	[redacted]	0.518	0.474	£11,866	£25,061
Analysis 2: Removal wastage for ruxolitinib	2.394	1.359	[redacted]	2.912	1.839	[redacted]	0.518	0.480	£13,958	£29,095
Analysis 3: Same AML rate (LT COMFORT-II)	2.394	1.362	[redacted]	2.912	1.836	[redacted]	0.518	0.474	£13,872	£29,243
Analysis 4: rate AEs from ITT population	2.394	1.359	[redacted]	2.912	1.839	[redacted]	0.518	0.479	£11,877	£24,776
ERG correction (analysis 1-4)	**2.394**	**1.359**	[redacted]	**2.912**	**1.827**	[redacted]	**0.518**	**0.468**	**£15,974**	**£34,147**
Scenario analysis: ERG correction + Gompertz for TTD (curves cross at 3.5 yrs)	2.065	1.233	[redacted]	2.548	1.606	[redacted]	0.482	0.372	£7,506	£20,160

7. Conclusion

In addition to the inputs/assumptions described in this document, there are a number of additional uncertain parameters/assumptions that are likely to affect the ICER, including (1) the proportion of patients on ruxolitinib in the BAT arm, (2) the proportion of patients remaining on JAK over time, (3) the cost for ruxolitinib, (4) time to death following discontinuation (extrapolation based on a single event for responders).

1. Celgene. **Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; Company's addendum submission to NICE**. 2021.
2. (EMA). European Medicines Agency. Assessment report: Inebric. 2020. (Updated: 10 December 2020) 2020. Available at: https://www.ema.europa.eu/en/documents/assessment-report/inrebic-epar-public-assessment-report_en.pdf. (Accessed March 2021).
3. Martyn-St James M Rafia R, Stevens J. Fedratinib for splenomegaly and symptoms in myelofibrosis. ERG commentary on the company's technical engagement response. *School of Health and Related Research (ScHARR)* 2020.
4. Celgene. **Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; Company's addendum submission to NICE - Clarification response**. 2021.
5. Martyn-St James M Rafia R, Stevens J, Rawdin A, Wong J. Fedratinib for splenomegaly and symptoms in myelofibrosis: A Single Technology Appraisal. *School of Health and Related Research (ScHARR)* 2020.
6. Celgene. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; Company's technical engagement response. 2020.
7. statement Clinical expert. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; Company's technical engagement response. 2020.
8. Celgene. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; 2020.

National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

ERG report – factual accuracy check and confidential information check

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

'*Data owners will be asked to check that confidential information is correctly marked in documents created by others in the technology appraisal process before release; for example, the technical report and ERG report.*' (Section 3.1.29, Guide to the processes of technology appraisals).

You are asked to check the ERG report to ensure there are no factual inaccuracies or errors in the marking of confidential information contained within it. The document should act as a method of detailing any inaccuracies found and how they should be corrected.

If you do identify any factual inaccuracies or errors in the marking of confidential information, you must inform NICE by **5pm** on **Friday 25 September 2020** using the below comments table.

All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

Please underline all confidential information, and separately highlight information that is submitted as 'commercial in confidence' in turquoise, all information submitted as 'academic in confidence' in yellow, and all information submitted as 'depersonalised data' in pink.

Issue 1 Clinical evidence

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 14. The key clinical evidence submitted by the company is derived from a single-arm, open-label, non-randomised, phase 2, multicentre study (JAKARTA-2); of fedratinib in myelofibrosis patients who were currently or previously treated with ruxolitinib for at least 28 days and who had symptomatic intermediate-1 risk, intermediate-2 or high-risk disease.	Please amend this statement to reflect the criteria for enrolment: patients had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib and therefore were not currently treated with ruxolitinib nor were they previously treated for at least 28 days.	Factual inaccuracy.	Report amended
Page 18. The clinical evidence provided in the CS comprises a single-arm study (JAKARTA-2) of fedratinib in myelofibrosis patients who were currently or previously treated with ruxolitinib.	Please amend this statement to reflect the criteria for enrolment: patients had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib and therefore were not currently treated with ruxolitinib.	Factual inaccuracy.	Report amended
Page 22. Appendix D.1.2 of the CS reports that the citation sifting stage was undertaken by a single reviewer but did not report if any secondary independent checking (considered systematic review best practice) of either all records or a proportion was undertaken.	Apologies if this was not clear in the original CS but please amend this statement to reflect the true approach taken to review which was that studies were assessed for eligibility by two independent reviewers, with disagreements adjudicated by a third reviewer. This applies to both primary (title/abstract) screening and secondary (full text) screening as confirmed in the response to clarification question C1 regarding this.	Factual inaccuracy.	Not a factual inaccuracy. As this detail was not reported in the original CS, this *post hoc* request cannot be validated by the ERG. Report unchanged
Page 24. Inclusion criteria for	Please amend this statement to reflect the fact	Factual inaccuracy.	Report amended

JAKARTA-2 were adults aged ≥18 years who had a current diagnosis of primary myelofibrosis, post-polycythaemia vera myelofibrosis, or post-essential thrombocythaemia myelofibrosis according to the 2008 WHO classifications, of intermediate-1, intermediate-2, or high-risk disease (according to the Dynamic International Prognostic Scoring System).	that risk categorisation was carried out using the IPSS or DIPSS in patients enrolled after Protocol Amendment 3 (see page 23 of the CS)		
Page 72. The key clinical evidence submitted by the company is derived from a single-arm, open-label, non-randomised, phase 2, multicentre study (JAKARTA-2); of fedratinib in myelofibrosis patients who were currently or previously treated with ruxolitinib for at least 28 days and who had symptomatic intermediate-1 risk, intermediate-2 or high-risk disease.	Please amend this statement to reflect the criteria for enrolment: patients had to have received ruxolitinib treatment for ≥ 14 days and have discontinued ruxolitinib for ≥ 14 days prior to receiving fedratinib and therefore were not currently treated with ruxolitinib nor were they previously treated for at least 28 days.	Factual inaccuracy.	Report amended

Issue 2 Including DIPSS criteria in the matching analyses

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 70 of the ERG report, the ERG states that "The CS states that risk categorisation was carried out using the International Prognostic Scoring System (IPSS) or the Dynamic International	The company suggests changing the wording of the paragraph to something like the following: The CS states that risk categorisation was carried out using the International Prognostic Scoring System (IPSS) or the Dynamic	The paragraph suggests that age, white blood cell count, haemoglobin, peripheral blood blasts and constitutional systems were not included in the matching analyses because they were not	Report amended

Prognostic Scoring System (DIPSS). Both categorisations includes age, white blood cell count, haemoglobin, peripheral blood blasts and constitutional systems. Hence, the ERG suggests that these variables (and not the resulting categories) should be included in any propensity score type model irrespective of their statistical significance. Absence of evidence that a variable is prognostic is not the same as evidence of absence of a variable being prognostic, and external clinical opinion should be used to guide which variable are included in or excluded from the model." The company believes that this paragraph does not give a true reflection of the process that was undertaken to identify prognostic factor/treatment effect modifiers.	International Prognostic Scoring System (DIPSS). Both categorisations include age, white blood cell count, haemoglobin, peripheral blood blasts and constitutional systems. The ERG suggests that these variables (and not the resulting categories) should be included in any propensity score type model irrespective of their statistical significance, where possible. SIMPLIFY-2 did not report white blood cell count, peripheral blood blasts and constitutional symptoms and therefore it was not possible to match on these three variables. The company did not match on age and haemoglobin due to the balance across studies. External clinical opinion should be used to guide which variable are included in or excluded from the model.	statistically significant in the univariable and multivariable analyses - which was not the case. As per the response to clarification question A8, white blood cell count, peripheral blood blasts and constitutional symptoms were not reported for SIMPLIFY-2 and could therefore not be matched on. Age and haemoglobin were not included due to the balance across studies.	

Issue 3 **Pooling of PERSIST-2 and SIMPLIFY-2 baseline characteristics**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 70 of the ERG report, the ERG states that "The CS considered variables for inclusion in the propensity score type model depending on whether there was	Suggest deleting this text.	The company did not pool characteristics of BAT-treated patients from SIMPLIFY-2 and PERSIST-2 because of the issues which the ERG outline.	No change.

an imbalance of ≥10 between JAKARTA-2 patients and the pooled BAT treated patients in PERSIST-2 and SIMPLIFY-2. The ERG does not consider it appropriate to simply pool the evidence on BAT because this ignores potential heterogeneity in the baseline response. Also, in Section 2.9.5.3 of the CS, the company wrote that, "PERSIST-2 is not representative of patients receiving ruxolitinib in BAT in the UK as it included patients with platelets <50 x 109 /L, for which ruxolitinib is not licensed." However, no pooling of PERSIST-2 and SIMPLIFY-2 characteristics was ever performed.		Standardized differences were estimated to investigate imbalance between JAKARTA-2 patients and SIMPLIFY-2 BAT patients only. These standardized differences were not used to select the matching variables. Variables were used in the matching if they satisfied both the following criteria: - The variable was identified as having clinically meaningful imbalance by an external hematologist The variable was also identified as being an important prognostic factor in the JAKARTA-2 study (from either the univariate or multivariable analyses)	

Issue 4 **Reason for not matching on mean haemoglobin and mean platelet count in analyses performed for A8**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 71, the ERG states that the additional MAICs performed in response to question A8 did not match on mean haemoglobin and mean platelet count because of "balance across studies".	Change reasoning to that described in the response to A8: "we were unable to match on mean haemoglobin and mean platelet count as baseline values for these variables were not available in the JAKARTA-2 patient-level data (only grouped variables)."	The company originally sought to adjust for baseline characteristics that were prognostic/treatment effect modifiers that were in imbalance across studies based on NICE DSU TSD 18: "Under this assumption, X must contain both every prognostic variable and every effect modifier that is in imbalance between the two studies". Following the suggestion	No change.

		from the ERG, the company attempted to adjust for all prognostic factors/treatment effect modifiers regardless of imbalance, however, were limited by the patient-level data available at the time.	

Issue 5 **Perspective stated within the CS**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
It is stated on page 80 of the ERG report that there is "no clear statement regarding the perspective", and on page 103 that the perspective was "not directly stated". The perspective for the model is stated in Table 1: The decision problem on page 10 of the report.	Remove sentence on page 80, change the "Perspective on costs" table entry on page 103	No impact on CE analysis, added as correction	This is not a factual error. The sentence on page 80 of the ERG report is taken out of context. The full sentence on page 80 is "*While there is no clear statement regarding the perspective, the company states that the reference case has been adhered to (CS, Table 1, page 9). Therefore, costs in the CS are considered from the perspective of the NHS and Personal Social Services (PSS)*." It is clear from this statement that this refer to costs. Similarly, it is clear that the statement on page 103 of the ERG report relates to costs as this in the row in the table discussing "*Perspective on costs*"

Issue 6 **ERG vague in issues with response rates used, potentially leading to incorrect interpretation**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG state on Page 108 that *response rates used in the economic model are unreliable*	The ERG should refer to the ITC and amend the sentence to state what about the ITC	Clarification of ERG concerns, removing scope for	The ERG agrees with the company that part of the statement is missing. The statement has been amended for clarity to read as follow.

because of significant differences between the population recruited in JAKARTA-2 and SIMPLIFY-2 Given that the response rates used are from the ITC, it is not correct in itself to state that the rates used are unreliable because the trials are different. To state this is to allude that the company has not conducted an ITC. To dismiss the response rate used because the trials are different questions the use of ITCs in general.	methodology makes them concerned that the response rates are unreliable	misinterpretation	"*response rates used in the economic model are unreliable because of significant differences between the populations recruited in JAKARTA-2 and SIMPLIFY-2 which have not been adjusted for in the unanchored ITC conducted by the company. SIMPLIFY-2 recruited patients who had myelofibrosis and previous ruxolitinib treatment for at least 28 days who either required red blood cell transfusions while on ruxolitinib or ruxolitinib dose reduction to less than 20 mg twice a day with at least one of grade 3 thrombocytopenia, anaemia, or bleeding at grade 3 or worse, with palpable spleen of at least 5 cm and without grade 2 or greater peripheral neuropathy. Inclusion criteria were different in JAKARTA-2. Not all of these differences in inclusion criteria were accounted for in the unanchored ITC conducted by the company. There are also issues described in Section 4.3.4.7*"

Issue 7 **The modelling approach of independent modelling of OS and TTD cannot be resolved robustly.**

Description of problem	**Description of proposed amendment**	**Justification for amendment**	**ERG response**
The company appreciate the ERGs feedback regarding the issues with independent modelling of OS and TTD. The ERG state on page 114 that *Given its modelling approach (independent modelling of OS and TTD), this issue cannot be resolved robustly* Whilst 'robustly' is a subjective term, it does	Remove 'cannot be resolved robustly' and replace the text by explaining that the issue cannot be resolved unless an assumption of dependence is applied in some	The amendment more accurately represents the setup of the model and the assumptions	This is not a factual error. The suggested approach is neither robust (due to discrepancies between TTD and OS), nor correct. Drawing from the same random number induces extreme dependence. While this makes the issue for the estimation of TTD less visible, it raises further questions about the validity and general implementation of the model. This is not how a DES should be implemented. TTD and OS are competing risks in the model and should be modelled as such. Alternatively, time to next

imply that there is no simple solution to model OS and TTD accurately in the current framework. A relationship between OS and TTD can be assumed if both events are sampled using the same random number to model events. It is acknowledged that there will be some discrepancy given TTD is estimated from week 24 and OS is estimated from week 0. However, overall, this aims to ensure that both TTD and OS curves are modelled according to their distribution. This is not a dissimilar approach to partitioned survival modelling. The sentiment is repeated on page 151, *"It should be noted that this cannot be avoided given the constraint imposed by the modelling approach where TTD is truncated by OS"*	way between OS and TTD.	used.	event should be modelled. References to the partitioned survival model to justify this approach is also surprising as in a partitioned survival model, the average distributions for the cohort is used and therefore there is no need to use random numbers to link the outcomes of heterogeneous individual's in this way.

Issue 8 **The value of separating responders and non-responders**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 114, the ERG state that *"the approach for estimating TTD was influenced by the approach to modelling (separating patients) rather than the other way around and therefore does not consider this argument to be valid"* This statement is not accurate. There was a clinical and conceptual	The ERG should acknowledge the conceptual expectation that responders would have a longer TTD than non-responders, and as such acknowledge the	The amendment more accurately reflects the choices made in relation to TTD.	This is not a factual error. The statement from the ERG is taken out of its context. As stated in the ERG report, the ERG does not consider the value of separating patients to outweigh consequences (section 3.4.3.2.3). In particular, the company does no use data for responders and non-responders using the same response definition. TTD could have been modelled as an average (which would account implicitly for differences in TTD between responders and non-

expectation that non-responders would discontinue sooner than responders, meaning that it was valuable to separate the TTD of responders and non-responders in the model. This expectation was supported by the findings of the JAKARTA-2 data and were shared with the ERG in the clarifications feedback question A1. The company have attempted to model the key model outcomes robustly and appropriately.	value in separating these patients.		responders) as is done for OS in the company model (which is modelled as an average for responders and non-responders including early death and discontinuation; despite these groups having different prognosis). The approach for TTD is therefore influenced by the company decision to split patient, rather than the other way around.

Issue 9 Incorrect values for non-responder numbers used

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 116: *"The ERG notes that the KM for non-responders provided by the company is not correct, as the KM includes [redacted] patients (n=[redacted] events, n=[redacted] censored) despite this subgroup based on only [redacted] patients (or [redacted] if using spleen definition)."* To assert that the KM provided is incorrect is not appropriate. The ERG's expected n of [redacted] relates to the number of non-responders (by spleen or symptom response) who were still receiving fedratinib at week 24 in JAKARTA-2 in the Int2/high risk population.	The ERG should remove the first part of the last paragraph on page 116. The responder and non-responder KM may be drawn. It should be highlighted that there were only [redacted] events from the responder population, which therefore gives	Factual inaccuracy to state that the provided data is not correct.	This is not a factual error. The statement is taken out of context and read in the ERG report: "*Additional clarification was sought from the company (see clarification response, question A1) to provide the KM for OS for responders (n=[redacted]) and non-responders (n=[redacted]) from 24 weeks. The ERG notes that the KM for non-responders provided by the company is not correct, as the KM includes [redacted] patients (n=[redacted] events, n=[redacted] censored) despite this subgroup based on only [redacted] patients (or [redacted] if using spleen definition)."* The ERG requested the following analysis (see additional

The provided KM shows the OS of non-responders from week 24 regardless of whether they were still receiving fedratinib at week 24, as this data is in line with the original request of the ERG. The KM provided with [redacted] events and [redacted] censors is correct.	reason for the company not to separate the JAKARTA-2 OS data by response.		clarification letter question A1) • *A1_6 "OS for non-responders (defined as spleen or symptoms) in JAKARTA-2 with Int2/high risk pop (n=[redacted]?) – from 24 weeks onward*". The N number is clear from the ERG request. It is also noted that the same request was made for TTD (A1_4) and the company did provide data with the correct N (n=[redacted]).

Issue 10 **Lack of face-validity for the scenario analysis assuming surrogacy between spleen and survival**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG state on page 140 that there is a *"Lack of face-validity for the scenario analysis assuming surrogacy between spleen and survival"* The company believes that it is not accurate to suggest that the surrogacy approach lacks face validity, because *"Assumptions surrounding this scenario were not assessed by the ERG further"* and the assumptions presented in this scenario do not reflect the surrogacy approach as a whole. The ERG only present results for the option in which survival for the first 24 weeks is set equal to BAT survival. This analysis option will therefore deviate from the trial KM for	The company acknowledge the uncertainty associated with this specific scenario. However, the ERG should also present a comparison between OS generated using the alternative surrogacy scenario and the CS base-case for completeness.	The incomplete presentation of results has led to an inaccurate portrayal of the surrogacy analysis. Inclusion of the analyses described would be useful for decision making.	This is not a factual error. The ICER for the surrogacy scenario that is described in the CS (key scenario analysis – Table 86) lack face validity. On page 179, the CS states: "*The top five scenarios are further summarised in Table 86 with a description and rationale for each scenario. The scenarios that result in the largest impact on the ICER are those that test the OS modelling method used and BAT composition assumptions. These scenarios are included because there was no head-to-head data between fedratinib and BAT that would have informed the relative OS between the two arms and the composition of BAT*".

fedratinib. On this basis, on page 141 of the ERG report, it is noted that *"The ERG considers that a complete rethink is required by the company, should the surrogacy OS scenario be considered for decision-making."* However, the pre-selected assumptions for surrogacy in the model use treatment-specific parametric curves for the first 24 weeks. The outputs of this analysis, which were reported in the model, have better face validity when comparing the output to the KM, although this has not been reported by the ERG.			On page 182 the company further state "*The scenarios showed that the assumptions surrounding the OS modelling assumptions and the composition of BAT had the most significant influence on the model outcomes*." The statement from the ERG is taken out of context, with the ERG assessment focusing on the surrogacy scenario that is described in CS (Table 86). The full statement from the ERG is: "*In addition to the sources available to quantify the relationship between response and OS (which is already uncertain), a number of options are included within the economic model to conceptually link inputs. The ERG believes that, should the analysis be robust and conceptually valid, results should be broadly consistent with each other. The company report results from key scenarios in Table 86 (CS, page 181) and estimate* [redacted] *compared with BAT when OS is estimated using surrogacy.* *The ERG replicated this analysis using 10,000 patients and plotted the predicted OS for fedratinib estimated from this scenario analysis and OS generated in the base-case (direct Gompertz fit). It can be seen (Figure 27) that predicted OS for fedratinib (red line) for the scenario analysis described in the CS (CS, Table 86, page 181) is inconsistent with both its base-case estimate (blue line), and the trial KM survival function (black line), raising significant doubt of its validity…* *Assumptions surrounding this scenario were not assessed by the ERG further because of its lack of face validity as shown in Figure 27.*"

Issue 11 **Consistent assumptions between treatment arms**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 4, the ERG state that *"assumptions need to be consistent between treatment arms".* This is a subjective matter unless supported by the clinical expectations that the use of ruxolitinib and fedratinib will be uniform. Assumptions should be consistent between treatment arms so long as this reflects the expected treatment pathway and/or patient experience of the treatments.	Removal of the sentence.	A subjective matter is reported as an a priori fact and is not consistent with the clinical advice received by the company	This is not a factual error. This is the ERG's view and is consistent with clinical advice received by the ERG (Section 4.3.4.4) that it is inappropriate to assume that patients on ruxolitinib/BAT remain on treatment for life (sub optimally) due to the absence of alternative treatment, but that patient initiated on fedratinib are allowed to discontinue early and receive HU subsequently. The statement is also taken out of context and read: "*Assumptions need to be consistent between treatment arms. As per assumption made for BAT/ruxolitinib (and argument advanced by the CS), the ERG considers that patients on fedratinib should continue treatment (fedratinib sub-optimally) beyond TTD"*

Issue 12 **Concerns with the implementation of AML as a health state**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 120, the ERG describes that their analyses to assess the implementation of AML as a health state were performed *"when the option to re-calculate*	Removal of paragraph 4 on page 120.	The ERG has not explained why a key and relevant model setting was disabled before producing	The ERG agrees with the company that part of the statement is missing and without context could be misleading. The statement has been amended to clarify why this was done and why the ERG does not consider AML to be implemented correctly. The paragraph has been amended to read as follow

survival is disabled". Having disabled a key model setting which aims to ensure greater accuracy of the time spent with AML, the ERG present over-estimated predictions of the mean time spent with AML. On this basis, it is concluded that *"the ERG does not consider that the company has implemented the AML health state correctly."* It is not accurate to report that the health state has been implemented incorrectly based on this analysis.		results which claim incorrect implementation. The recalculation of survival is an important component of the analyses, and should be enabled if the ERG are to fully discuss the appropriateness of this approach.	"*Perhaps more importantly, as previously described in this ERG report, there are some programming errors in the CS.* *In the company base-case, survival for patients with AML is re-estimated using a formula described in Appendix L.8 of the CS. The ERG notes that when running the company's base case model for 10,000 patients, the model predicts that [redacted] patients in the fedratinib arm would transform to AML, of whom [redacted] would have a predicted time in AML of less than zero, which is not possible. Patients who are predicted to experience AML, with a time less than 8 weeks are also assumed to move directly to palliative care and therefore have a time in health state equal to zero. The time in the AML health state is estimated to be zero for [redacted] patients ([redacted]%) with AML (while the ERG understands this was intentional, this highlight some conceptual issues to the ERG).* *Recalculating OS also led to different OS prediction compared with the initial parametric fit. If the option to re-calculate survival is disabled (which is rightly not presented by the company), the time predicted in the AML health state would be over-inflated to a mean time of [redacted] weeks (median: [redacted] weeks), which does not align with the survival estimate from Mesa et al (2005) presented by the company in Appendix L.8 (median survival of less than 12 weeks). Approximately [redacted]% of patients with AML (n=[redacted]) would have a predicted time in this health state of more than 10 years.* *The company therefore has to recalculate survival, but this is not done correctly and a large number of individuals have a time in AML equal to zero* *Consequently, the ERG does not consider that the company has implemented the AML health state correctly as it generates negative values, a large proportion of patients with AML with a time in health state equal to zero and the estimated survival (after recalculation) do not match the initial OS distribution (as it is adjusted)*."

Issue 13 Concerns regarding the OS survival function selection process for patients initiated on fedratinib

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
It is appreciated that the CS and advisory board report potentially lacks clarity	The ERG should	The ERG have	The ERG appreciates the

on the issue of the OS survival function selection process, and this has led to factual inaccuracies in the ERG report. For clarity, the advisory board was approached with the OS KM for the ITT population, which is why the advisory board report only includes the ITT KM and extrapolations. During the advisory board, the company was informed that ITT was not appropriate for the UK population, and that the intermediate-1 patients (16%) should be excluded. In figure 2 of the ad-board report, the clinicians agreed that for the ITT curves, that exponential and Weibull distributions were reasonable, and was why Weibull was selected as the ITT base case. The clinicians gave the 'consensus values' on what the intermediate-2/high-risk population would be expected to look like. The clinicians agreed that for this population they would expect a curve of a similar shape to the preferred ITT curves selected, but poorer outcomes to account for the higher risk group. The closest curve to the consensus values was the Gompertz. Following the advice given at the advisory board, the intermediate-2/high-risk subgroup was assessed and extrapolated. It was found that the Gompertz curve was the closest curve to the consensus estimates, therefore it was used in the base case. Given this, there were factual inaccuracies in the ERG report: On page 123 the ERG state *"The approach taken by the company is also inconsistent as OS for patients with intermediate-2/high risk is likely to be worse compared with the ITT population"* This is not accurate as in the company base case assumptions, the OS for patients with intermediate-2/high risk MF <u>is worse</u> than that of the ITT population. On page 123, the ERG also state *"the ERG notes that clinical experts to the company considered that while the exponential and Weibull distributions "seems" reasonable, that the generalised gamma (not the Gompertz) is more clinically reasonable (for the ITT population)"* There are two inaccuracies in the statement above. It is noted on page 127 of the CS that *"For fedratinib, the clinicians were shown parametric curves fit to the*	correct or remove the text which describes potential inconsistencies.	misinterpreted, potentially due to a lack of clarity in the CS, the extrapolations provided by the UK clinical experts at the advisory board. This means that statements made by the ERG are incorrect and require amendment to ensure accurate reporting on the OS survival function selection process.	company acknowledging the lack of clarity in the CS and the company's clarification response. This section (4.3.4.5) was therefore amended to reflect this new information. However; the additional details provided by the company raised further questions. See Section 4.3.4.5 in ERG report for amended section.

survival data for the JAKARTA-2 ITT population (N=97). During the meeting it was advised that only the intermediate-2 and high-risk population would receive fedratinib in the UK; therefore, expectations of survival were provided with this in mind". Therefore, the first inaccuracy in the above statement is that the ERG write *"for the ITT population"*. The curve was selected as more clinically reasonable for the intermediate-2 and high-risk population, as described earlier. The second inaccuracy is that the final advisory board report replaces generalised gamma with Gompertz (which were highly similar in the curves presented to the experts). The reason for the change between versions is an error in the note-taking of the original author, that was misaligned with the notes taken by other attendees and therefore corrected. This misinterpretation, potentially due to a lack a clarity in the CS, is repeated on page 124: *"According to the curve selection process used by the company, the ERG would therefore expect the model for the subset of patients with intermediate-2/high risk to generate an OS estimate either similar to or lower than the ITT population. However, this is not the case. The model selected by the company for the subset of patients with intermediate-2/high risk (Gompertz distribution) predicts a more favourable survival compared with the ITT population (Figure 22) using the Gompertz distribution selected for the ITT population by its own clinical experts."* As described above, the clinicians provided guidance on survival expectations with the UK intermediate-2 and high-risk population in mind. As such, the inconsistency reported by the ERG is not accurate.			

Issue 14 Choice of survival function for Schain et al (2019)

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
On page 130, the ERG state that "*The company fitted parametric models to the BAT data from Schain et al (2019) and selected the Weibull survival function as the*	The ERG should update the text to	Factual inaccuracy.	We agree with the company and amended the statement to reflect both the CS and advisory board notes:

base case because "it provided a better statistical fit to the data." The ERG notes that the model with the smallest BIC was the lognormal distribution." At the clarification stage, in response to the related clarification question (A21), it was noted that this quote is taken out of context, and that reading the preceding sentence shows that the comparison is made to the fit for the exponential model (the other distribution deemed clinically plausible by the clinical experts). In the CS, page 125, clinical input is also reported: *"the group indicated that the exponential and Weibull were most relevant and representative of UK patients. The Weibull curve was selected in the base case as it provided a better statistical fit to the data"* The above text reported by the ERG suggests that only statistical fit was considered which is not accurate.	accurately describe the reason why the Weibull survival function was selected.		*"The company fitted parametric models to the BAT data from Schain et al (2019) and selected the Weibull survival function. The CS stated (page 125) that "the group indicated that the exponential and Weibull were most relevant and representative of UK patients. The Weibull curve was selected in the base case as it provided a better statistical fit to the data". The ERG notes the following statement in the advisory board notes that "Exponential & Weibull extrapolations are the most relevant to the dataset (see figure 1). Exponential may better represent the survival to UK patients." While the Weibull distribution provide better statistical fit to the data, the exponential distribution (appeared to be preferred by clinical experts at the advisory board) and the Weibull distribution provided very different long -term extrapolation"*.

Issue 15 ERG comparison between JAKARTA-2 and SIMPLIFY-2 surmised incorrectly

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 133 states that patients in the SIMPLIFY-2 study have a higher platelet count versus JAKARTA-2.	Page 133 paragraph 2 should be changed to state that the SIMPLIFY-2 platelet count is lower than JAKARTA-2	The values given by the ERG for the platelet count in both studies are correct, however the SIMPLIFY-2 values are lower. Minor correction. No impact on the interpretation or overall report	The ERG agrees with the company – the text has been amended as requested

Issue 16 **ERG stating that the mean platelet count of SIMPLFY-2 was 126.5 x 10^9/L**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG claim on page 144 that the mean platelet count for SIMPLIFY-2 was 126.5 x 10^9/L. This is incorrect, as the number reported is the value for the SIMPLIFY-2 BAT arm. The SIMPLIFY-2 momelotinib arm has a platelet count of 178.2 x 10^9/L, therefore the true mean platelet count of SIMPLIFY-2 is between these two values	Within the cost arguments in section 4.3.4.11, state that the platelet count value of 126.5 x 10^9/L is only associated with the BAT arm of SIMPLIFY-2, and include the momelotinib arm has a platelet count of 178.2 x 10^9/L in the text. It is appropriate to use the platelet count value of 126.5 x 10^9/L for the text on page 133 regarding the ITC because only the BAT arm was used for this, however clarifying that the values quoted only correspond to the BAT arm would make this clearer.	The mean platelet count for JAKARTA-2 was [redacted] x 10^9/L. Patients with a platelet count < 100 x 10^9/L receive low dose ruxolitinib (5mg BID) and patients with a platelet count ≥ 100 x 10^9/L receiving higher ruxolitinib dosages. The ERG argue that because of the difference between 126.5 x 10^9/L platelet count mean in SIMPLIFY-2 and the JAKARTA-2 platelet count mean that the costs are overestimated for ruxolitinib in BAT in the model comparator because it is not reflective of the full population of interest. Whilst it is acknowledged that there is a difference, it is misleading of the ERG to only quote the lower value of the two randomized SIMPLIFY-2 populations, as the higher value from the randomized trial is far more similar to the values used in JAKARTA-2.	The statement has been amended to clarify that the value quoted is for the BAT arm of SIMPLIFY-2 (where efficacy is taken from), and not the whole SIMPLIFY-2 population. The platelet count in the momelatinib arm of SIMPLIFY-2 is not relevant as this arm is not used by the company. Consequently, the sentence has been amended as follow: "*While the distribution of platelet count (patients with a platelet count < 100 x 10^9/L) is not reported in SIMPLIFY-2, the mean platelet count was 126.5 (SD: 95.9) x 10^9/L for BAT in SIMPLIFY-2 (arm from which efficacy [response rate] is taken from for the comparator arm) vs. [redacted] (SD: [redacted]) x 10^9/L in JAKARTA-2. Although the magnitude of bias is uncertain without access to data from SIMPLIFY-2 the ERG believes that the cost for ruxolitinib to be overestimated by the company and not aligned with the platelet count distribution for the BAT arm in SIMPLIFY-2 (from which efficacy data [response rate] is taken from)*".

Issue 17 **The ERG claimed that the CS base-case does not include adverse event impact on quality of life**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 147 states "The CS base-case	Change the sentence to state	Factual inaccuracy, minimal	The ERG agrees with the company – the text

does not include the impact of AE on quality of life; this is presented as a scenario analysis only". This is incorrect as both the original model and the model delivered to the ERG as part of the clarification questions included AE disutility within the base case (Controls sheet "c_include_disutility_text")	that AE disutility was included in the CS base case	model impact	has been amended as requested and now read as follow: "*The CS base-case includes the impact of AE on quality of life separately. The ERG does not consider this to be appropriate as utility values are taken from JAKARTA-2 and already include the effect of adverse events associated with fedratinib. The ERG further notes that decrement in utility values are taken from a range of sources using varying preference based measures, and therefore there are concern with mixing values from the MF-8D with other measures. However, this is a small issue*."

Issue 18 The ERG claim that the end of life criteria of survival of less than 24 months is not met

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG state the following as part of issue 11: *The company's base-case predicts a mean LY over 2 years ([redacted] years), and therefore does not meet EoL criteria* This is factually incorrect as the NICE criteria for this point states: *The treatment is indicated for patients with a short life expectancy, normally less than 24 months* This does not specify that there has	The point that end-of-life criteria is not met solely because of model mean LYs should be amended to acknowledge the median LYs in the company base case.	Factual inaccuracy and misinterpretation of NICE end of life criteria. Median LYs associated with BAT is 1.201	This is not a factual error. This sentence is taken out of context. The beginning of the section clearly highlights the NICE EoL criteria. It is then clear that it is the ERG's view (not NICE) that the model predictions do not meet this criterion. The statement in the ERG report reads as follows: "*the ERG consider it to be debatable whether the population entering the economic model (as defined by the company – patients on suboptimal ruxolitinib) meet the EoL criteria for the following reasons:* • *The EoL criteria is not met in the company's economic model (for its base-case) as it predicts a mean LY for BAT over 2 years ([redacted] years). It should be noted that this increases even further under the ERG*

to be a mean LY of less than 24 months. Mean LYs can be inflated when using extrapolations with long tails. Therefore, the claim on its own that because the mean LYs is over 24 months means that the EOL criteria is not met is factually incorrect. In the CS base case the median LYs associated with BAT is 1.201.			*preferred base-case*" The median is also irrelevant in the context of an economic evaluation. However, the median survival (from Schain used in the economic model and other selected studies by the company) has already been already acknowledged in the following statement "*The CS states that these studies show a median OS of 13–16 months and these estimates are likely to be even lower in the intermediate-2 and high-risk population*".

Issue 19 Misleading presentation of results

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 1.5, Page 2, Issue 4, it is stated that *"when TTD is corrected the ICER change from being dominant to £444,999 per QALY gained"*. This large jump is framed as being due to correcting TTD, but this is misleading as it reflects assumptions related to Issue 7. For Table 2 and Table 33, the order in which issues are presented *(i.e. the analysis which assumes no survival benefit is presented before 'correction of errors')* exaggerates the extent of the impact of the original TTD approach.	The report should detail the extent of the ICER impact with and without the assumption of no survival difference to fully present the impact of the change in TTD alone. When the ERG's correction to TTD is applied to Analysis 1 + 2, the ICER changes from £8,303 to £43,106. The ERG should present analyses in which no survival difference is assumed in the bottom rows of Tables 2 and 33, to demonstrate that the drastic changes in the ICER are driven by this assumption.	The key driver of the high ICERs preferred by the ERG are driven by the assumption of no survival difference between fedratinib and BAT. The 'correction of errors' related to time on treatment will increase the ICER but to a much lesser extent to that which is reported. This should be made clear to better inform decision making.	This is not a factual error. The ERG's exploratory analyses are intended to address key issues in the company's model. The ERG does not believe that there is any evidence to demonstrate any survival gain for fedratinib.

Issue 20 'Full rethink' terminology and approach

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG uses the phrase 'full' or 'complete' 'rethink' 14 times in the report to describe their recommended approach to the logic used. An example of the context is used below: *These conceptual and structural problems are complex and intertwined, and the resolution of individual issues in isolation would not result in an appropriate or credible model. Rather, the ERG considers that the joint resolution of these problems would require a 'full' rethinking of the model's logic* By stating that the issues are complex and intertwined, and then relying on that assumption, the ERG have presented thinking into the individual issues they highlight, thereby overlooking (i) how resolving one leads to the resolution of others, (ii) the reasoning for the approach by the company, (iii) how the approach is applied in the model and the	Remove the term full 'rethink' where it relates to factual inaccuracies highlighted in this document. Justify the reasoning for the ERG believing a full rethink is required where there are issues remaining.	Unfair representation of the model to the extent where there is serious concern that the language would influence decision making.	This is not a factual error. It is the ERG's assessment that the model as submitted by the company is not fit for purpose and requires a full/complete rethink given its current structure. This is illustrated further by the response from the company to issue 7 resulting in additional concern regarding the general modelling approach and implementation. None of the points raised at factual accuracy check changed the ERG assessment.

various options available, or (iv) steps that could be taken to lessen the severity of particular issues. This is an oversight leading to the model being presented as entirely unfit, and is linked to many of the factual inaccuracies that are outlined in this document. The main issues the ERG associate with a 'full rethinking' are presented in Table 1, below.			

Issue 21 'Conceptual errors' terminology

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Some of the issues described in Table 1 and Issue 20 are referred to as 'conceptual errors', and are grouped together indiscriminately. The term 'error' is considered to be objective, however we are concerned that it has not been used this way throughout the report. We understand that the ERG may disagree with some of the assumptions and conceptual approaches used within the model, however it is misleading to	Unbiased language is preferred such as 'conceptual differences'.	Unfair representation of the model to the extent where there is serious concern that the language would influence decision making.	This is not a factual error. The term "*conceptual errors*" is used in 3 places in the ERG report: • P106: "*However, during the model verification process (in particular checking predictions from the model), the ERG identified a number of programming/conceptual errors in the implementation of both the AML and palliative care health state*" • Page 110: "*However, the current implementation in the CS introduces a number of biases, and mathematical and conceptual errors, with results for relevant scenarios (such as examination of the impact of the stopping rule, estimation of OS through surrogacy relationship and including the worsening in quality of life) lacking face-validity*". • P117: "*Following review of the CS model predictions, the ERG identified a number of programming and conceptual errors, some identified at the*

call these errors without describing exactly how the current approach is made in error, as opposed to being a matter of subjective opinion.			*clarification stage*" It is clear from the placement of this statement by the ERG and the context of sentences containing this statement that these relate to the implementation of the AML health state and the general modelling approach. It is the ERG 's assessment that the company's model is not implemented correctly and does not predict its own data, generates negative values and zero time in health state (for AML) and predicts inappropriate values for time in health states. The ERG consider those to be conceptual errors; not conceptual differences.

Table 1: Table of issues identified by the ERG associated with a 'full or complete rethink' required

Issue presented by ERG justifying a 'full rethink'	Reference in ERG report	Company response
Inaccurate estimation of TTD for fedratinib and introduction of unnecessary uncertainties	Section 1.5, Issue 4 Section 4.3.4.2.2 Section 4.3.4.2.3	It is appreciated by the ERG that the CS assuming independence between TTD and OS presented an inaccurate estimation of TTD for fedratinib. As described in Issue 7 of this document, dependency can be incorporated into the existing model without undertaking a 'full rethink'. Implementing such an approach removes some of the uncertainties that the ERG associate with the independence assumption.
Conceptual inconsistencies that responders have the same survival as non-responders	Section 1.5, Issue 4 Section 4.3.4.2.4.	As outlined in Issue 9 of this document, the JAKARTA-2 OS data was intentionally not separated by response by the company due to limited data.
Inaccurate prediction for the AML health state	Section 1.5, Issue 4 4.3.4.3.2	A factual accuracy response has been provided in Issue 12 of this document.
Omission of supportive care health state	Section 1.5, Issue 5 4.3.4.3.1	Not a factual inaccuracy. However, this point can be addressed by the company, and the point alone does not warrant the terminology of a 'full rethink' of the model.
HRQoL in patients initiated on fedratinib	Section 1.5, Issue 5 4.3.4.13	Not a factual inaccuracy. However, the point is viewed as a minor issue that can be addressed by the company, and the point alone does not warrant the terminology of a 'full rethink' of the model.
Inconsistent assumption between BAT and fedratinib	Section 1.5, Issue 6	A response by the ERG to this point has been provided in Issue 11 of this document. The model was also made sufficiently flexible during the clarification stage by the company so that the ERG could explore the impact of including fedratinib in the post-fedratinib BAT arm.
Lack of face-validity for the scenario analysis assuming surrogacy between spleen and survival	4.3.4.9	A response by the ERG to this point has been provided in Issue 10 of this document.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments: **5pm on Thursday 5 November 2020**.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.

- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Celgene, a BMS Company**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: Concerns with phase 2, single-arm JAKARTA-2 study	**NO**	The single-arm design of JAKARTA-2 is considered justifiable for a population with a high unmet need, such as those who have failed treatment with ruxolitinib. These patients have no other treatment options. As such, this trial is also informing the market authorization application for fedratinib in Europe. Understandably, there is uncertainty associated with use of data from JAKARTA-2. Celgene considers fedratinib to be a candidate for entry into the CDF, so that the uncertainty can be addressed without delaying patient access to treatment. Impending clinical evidence to support routine use of fedratinib includes FREEDOM-2, which is anticipated to read-out in 2022. FREEDOM-2 is a Phase 3, open-label, randomised study designed to evaluate the efficacy and safety of fedratinib compared to BAT in patients with intermediate or high-risk primary myelofibrosis, post-polycythemia vera myelofibrosis, or post-essential thrombocythemia myelofibrosis and previously treated with ruxolitinib. Examples of other treatments recommended for CDF funding based on single-arm data include TA567, TA554, TA559, TA510 and TA592.
Key issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT	**YES**	*Scale of the relative treatment effects* The original model presented estimates of relative treatment effect on the absolute risk scale (as risk differences). Relative treatment effects have been updated in the

		model to be on the additive scale (as odds ratios). This also avoids the need to cap response rates in the cost-effectiveness model. *Prognostic factors and treatment effect modifiers adjusted for in the MAICs* Given the relatively small sample size of the JAKARTA-2 study, a pragmatic approach was taken to identify the most important baseline characteristics in imbalance, based on those available across the two studies (JAKARTA-2 & SIMPLIFY 2). Variables were used in the matching if they satisfied the following criteria: 1) The variable was identified as having clinically meaningful imbalance by an external haematologist and 2) The variable was also identified as being an important prognostic factor in the JAKARTA-2 study (from either the univariate or multivariable analyses). The haematologist identified ECOG PS, DIPSS and transfusion dependence has having a clinically meaningful difference between the two studies. Variables in imbalance were used in the matching based on the NICE DSU TSD 18 guidance ("X must contain both every prognostic variable and every effect modifier that is in imbalance between the two studies – an assumption that is largely deemed unreasonable") which is supported by the recent simulation study by Hatswell et al. 2020 ("The Effects of Model Misspecification in Unanchored Matching-Adjusted Indirect Comparison: Results of a Simulation Study") which found that the performance of matching methods worsened when variables were included that were already well matched between studies or that were not linked to outcomes. In the base case analysis, the effective sample size was 81 patients, which was deemed sufficient in this context. Important variables in imbalance were also included in the matching as sensitivity analyses that included more variables. This heavily reduced the effective sample size indicating that estimates are unstable and inferences depend heavily on just a small number of individuals e.g. the effective sample size was 15 when adjusting for ECOG PS, DIPSS, transfusion

		dependence and age. It was considered not appropriate to draw conclusions from analyses with a small effective sample size (like for a small sample size). See Table 1 Appendix
Key issue 3: Alignment between the comparator and the modelled population	**YES**	The ERG report stated that the efficacy inputs were not aligned with the population entering into the model because the data is taken from patients who are no longer treated with ruxolitinib. The ERG suggested to separate the patients to (1) those who switch to fedratinib at the point of resistance/refractory/intolerance (r/r/i) and (2) patients who discontinue ruxolitinib. It was not possible to split the JAKARTA-2 population as suggested given that there is considerable overlap between these populations that the ERG describe. For the BAT arm, further data was sought to inform the OS of patients from the point of r/r/i. Data was accessed from an ongoing global chart review which includes patients from the UK being conducted by Celgene, a BMS company. The chart review aims to investigate the management of Myelofibrosis patients which included survival, in a population more closely aligned to that entering JAKARTA-2. These data were added to the updated cost-effectiveness model as a means to address uncertainty regarding the appropriate OS baseline for BAT. See Figures, 1, 2 3 & 4 in Appendix
Key issue 4: Inappropriate approach to modelling	**YES**	The approach to model TTD and OS in the original submitted model derived the timing of both events as independent of one another. Therefore, the final time on fedratinib treatment could be lower than the sampled time if time-to-death was sampled before the time-to-discontinuation. The resulting time on treatment from the total patient cohort had a median value that was less than the extrapolated data. The ERG was concerned with this approach and it was agreed that the

		model should include the functionality to allow interdependence between TTD and OS. In response to the ERG report, structural changes were made to the model to allow for such analysis. Firstly, TTD is now estimated from model entry, rather than using inputs split pre- and post- 24 weeks. For OS, two options have been included that introduce dependency between OS and TTD. One approach estimates fedratinib OS from the point of discontinuation, and the second approach uses a common random number to derive both TTD and OS events. In addition, the OS and TTD outcomes for both the fedratinib arm and the BAT arm are modelled according to 24-week response outcome. The ERG described how the approach to modelling meant that corrections to TTD (under ERG preferred assumptions) would result in the ICER for fedratinib moving from dominant to over £400,000 per QALY. The company would like to stress that the ICER is highly sensitive in this scenario because the ERG assumed no survival gain on fedratinib. The ERG raised a separate concern regarding the implementation of AML as a health state. As part of this, the ERG did not agree that time-to-death should be re-estimated upon AML health state entry. Celgene still believe this approach is valid, however, in response to the ERG, and as part of the suite of changes implemented to better align with the ERG's preferred approach, the AML health state was removed, and AML was instead implemented as an adverse event. We believe that these amendments to the model constitute an overall improved analysis, with the goal of resolving the stated concerns of the ERG surrounding the structure of the model. See Figures 5, 6, 7, 8, 9 & 10 in Appendix
Key issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	**YES**	Based on clinical advice Celgene, a BMS company have received, the original submitted model assumed that patients remained on treatment for life in the case of the best available therapy (BAT) arm (with the exception of patients transforming to AML). It also assumed that patients receiving BAT as post-

		fedratinib therapy would also remain on BAT until end of life. This approach was not considered appropriate by the ERG, who thought it more appropriate for patients to discontinue to supportive care before death, owing to the data from HMRN and the approach taken in TA386. The model was updated in response to the ERG report by including a supportive care health state that uses the same resource use assumptions as were applied in TA386. The HRQoL assumptions for supportive care were also taken from TA386. Data from HMRN (2020) was used to estimate TTD for the BAT arm, whereas post-fedratinib transitions to BAT, supportive care and death were informed by inputs from clinical opinion. The ERG also highlighted that patients moving from fedratinib to BAT were modelled to experience HRQoL on BAT that was similar to patients treated with a JAK inhibitor (albeit still based on whether the patient responded) and felt that this was not plausible. Two changes to the model were made to account for this: • Firstly, the model was updated such that patients would not undergo a second 'response assessment' after discontinuing fedratinib. Therefore, patients could not receive a responder increment in utility beyond discontinuation. • Secondly, since the utility data in the model is derived from JAKARTA-2 which was a single-arm study, the utility increment for 'non-responders' is reflective of a group of patients who did not meet the dichotomous responder criteria, but still achieved some clinical benefit.[1] In contrast, non-responding patients receiving BAT in SIMPLIFY-2 often had increases in spleen size.[2] Therefore, it was felt plausible that non-responders to BAT (and patients receiving BAT after fedratinib) should have no increment in utility. See Table 2 and Figure 11 in Appendix

Key issue 6: Inconsistent assumption between BAT and fedratinib	**YES**	It was acknowledged that there was a difference between how the time on treatment was applied between the BAT and fedratinib arms. In response, the updated model structure and associated assumptions (e.g. time spent in supportive care) as outlined in our response to Key issues 4 and 5 will lead to time on treatment that is more in line with ERG expectations than the original submitted model. It should be noted however that differences are reported between the two treatment arms and it is appropriate to model them where there is supporting evidence. The continuation of suboptimal ruxolitinib is part of the current clinical practice in the UK confirmed by clinical experts and has been acknowledged by the ERG. Therefore, the application of suboptimal ruxolitinib model is an informed clinical pathway as opposed to an assumption. Additionally, the FREEDOM2 study is likely to have a high proportion of ruxolitinib within the BAT arm. On the other hand, patients continuing fedratinib is an assumption as there is no evidence of their continuation. The TTD curve derived from JAKARTA-2 event data captures the length of time patient are treated with fedratinib.
Key issue 7: Assumption of survival difference	**YES**	It is understood that JAKARTA-2 is a single arm trial without a direct head-to-head comparator. It is therefore difficult to conclusively prove a survival difference. We attempted to address this in the original submission through completing a feasibility assessment for an indirect treatment comparison for overall survival, however none of the available data sources fulfilled the requirements for completing an ITC. Therefore, the model has relied on naïve comparisons. The model updated in response to the ERG includes 7 different OS sources for BAT, which vary in appropriateness for the population of interest; some data sources include intermediate-1 patients and others include patients where a vast proportion

		receive ruxolitinib. In all of the presented cases, fedratinib demonstrates survival benefit at all time points of the available data. When the ERG compare survival of refractory patients in JAKARTA-2 to that in Miller et al (after 3 months of ruxolitinib treatment), and Palandri et al (after 6 months of ruxolitinib), making the suggestion that there is no survival gain, this comparison fails to highlight that patients in JAKARTA-2 had a median duration of ruxolitinib exposure of [redacted] months. In Schain et al., the source informing the base case OS in the original submission, patients discontinued ruxolitinib after a similar median treatment duration of 11.5 months. OS data for BAT in SIMPLIFY-2 was presented in a slide deck developed by Sierra Oncology. As OS was not a specified endpoint in the SIMPLIFY-2 trial, this was a post-hoc analysis. It was assumed initially that the data would be reported correctly by the authors. However, due to discrepancies within the source itself, further assessment was conducted, and the accuracy of this evidence is uncertain given the following: • The slide deck was found through a Google search and the OS evidence has not been presented in an accredited, peer-reviewed journal • The text describing OS in SIMPLIFY-2 refers to a double-blind treatment phase but SIMPLIFY-2 was open-label, casting doubt on the accuracy of the data reported • Assuming the author of the slide deck meant 24 weeks when referring to the double-blind treatment phase, the text describing OS at the end of 24 weeks does not match the Kaplan-Meier presented: o The text stated that 21% of BAT subjects had died after 24 weeks, whereas the pseudo Kaplan-Meier data indicates that 4 of the 52 patients had died during the first 24 weeks (8%) Further, as patients in the BAT arm of SIMPLIFY-2 could receive momelotinib after 24 weeks, the data after 24 weeks is not representative of patients receiving BAT.

		The pseudo patient-level data was therefore censored at the end of Week 24. 24 weeks is not considered a long enough time to assess overall survival data. A benefit of fedratinib is demonstrated by the proportion of patients who achieve a spleen, symptom or 'spleen or symptom' (Int-2/HR population: 33.3%, 28.4% and 49.4% response respectively) response. There was a significant difference reported in the proportion of patients achieving a response in the ITCs to both PERSIST-2 and SIMPLIFY-2. Supportive evidence for a relationship between spleen response and overall survival has been presented in the original submission.[3-5] The Miller et al. (2017) study highlighted by the ERG also provides supportive evidence of a link between spleen size reduction and survival. See Table 3 and Figures 12, 13 & 14 In Appendix.
Key issue 8: Lack of face validity for the stopping rule scenario	**YES**	The ERG reported their concern that a stopping rule in the economic model should impact OS. The new evidence provided in response to Key issue 4 will mean that scenarios with a stopping rule would impact OS, since OS for fedratinib could be estimated from the time of discontinuation.
Key issue 9: Costs for the comparator arm (ruxolitinib)	**YES**	Ruxolitinib is costed based on the platelet count distribution in JAKARTA-2, which is used to calculate the proportion of patients below and above the platelet count threshold of 100,000 x 10^9/ L. Patients above this threshold receive higher amounts of ruxolitinib which in turn accrues higher cost. It was noted by the ERG that the mean platelet count in JAKARTA-2 was higher than both the treatment arms in SIMPLIFY-2 and other studies, therefore there was concern that the costing of ruxolitinib would not be representative of the UK MF population. The SIMPLIFY-2 study did not publish the proportion of patients

		with a platelet count below 100,000x 10^9/ L but did publish the mean and standard distribution. Distributional assumptions using this data were made to estimate a proportion of patients receiving lower dose ruxolitinib that may be more reflective of the efficacy data in SIMPLIFY-2. These have been added as options to use within the model. In addition, recently published evidence from SIMPLIFY-2 from Gupta et al.[6] has noted that 86.7% of prior JAKi-treated patients commenced ruxolitinib treatment with a mean daily dose below the maximum ruxolitinib dose, including 26.7% receiving 5 mg BID or less. This suggests firstly that the currently modelled distribution of 34% receiving the lower dose and therefore cost of ruxolitinib may be *underestimated*, and that if this was considered then the ICER for fedratinib may be lower. Secondly, it reaffirms the clinical backing that patients currently receive suboptimal ruxolitinib treatment which would not be expected to significantly extend survival. See Table 4 in Appendix
Key issue 10: Reliability of response rate	**YES**	*Comparability of the population included in SIMPLIFY-2 and JAKARTA-2* A noteworthy difference between the JAKARTA-2 and SIMPLIFY-2 studies is the inclusion/exclusion criteria for prior ruxolitinib treatment. Subjects in SIMPLIFY-2 had not necessarily discontinued ruxolitinib on enrolment into the study and the inclusion criteria (subjects to be currently or previously treated with ruxolitinib and have either required red blood cell transfusion while on ruxolitinib, or had a dose adjustment of ruxolitinib to < 20 mg twice daily and Grade 3 thrombocytopenia/anaemia/hematoma) may have resulted in a high proportion of subjects who are transfusion dependent enrolled into the study (52% of BAT treated subjects in SIMPLIFY 2 are transfusion dependent compared with 14.4% in JAKARTA-2). An attempt was made to adjust for transfusion dependence in the MAIC analyses to account for this difference across studies.

		Comparability of study design which is likely to be more favourable to fedratinib (no washout period in SIMPLIFY-2) The response seen in the BAT arm of SIMPLIFY 2 is not likely to be different even if there had been a short washout of 14 days as these patients were already on ruxolitinib. Therefore, restarting them back on ruxolitinib is not likely to improve their existing response. *Estimation of the treatment effect using differences in absolute risk* As described in Issue 2, relative treatment effects have been updated in the model to be on the additive scale (as odds ratios). *Assumption that response rate for BAT after fedratinib is the same as BAT* The original submission assumes that post-fedratinib BAT did not include and JAK-inhibitors, therefore the ERG were concerned that the response rate applied post-fedratinib to patients receiving BAT would not be equal to the response in the BAT arm. As described in the response to Key Issue 5, the model was updated such that patients would not undergo a second 'response assessment' after discontinuing fedratinib. *Use of spleen response by volume rather than spleen length* It is known that the measurement of the spleen size has usually been done by palpation despite the irregular shape of the spleen. Measurement errors can arise due to subjective grading and variability of clinician's methods. Spleen Volume measurement by CT/MRI scan is seen to be more accurate and objective method to identify splenomegaly and is the primary endpoint for Myelofibrosis clinical trials. In the Miller *et al*[7] paper, patients treated with ruxolitinib who experienced worsening spleen length assessed using palpation at week 12 had a reduction in splenomegaly on the basis of spleen volume measurement and this was maintained over time. Even patients who achieved a minimal or no improvement in spleen length at week 12 while on ruxolitinib treatment, showed evidence of an improvement in spleen volume, MF-related symptoms, body weight and serum albumin levels. Spleen length alone was deemed insufficient to identify all patients

		benefiting from treatment. Spleen Volume measurement by MRI/CT scan enables clinicians to identify all such patients.
Key issue 11: End of life criteria	**YES**	The base case model including the changes made in response to the technical engagement show that the median LYs for patients on the BAT arm was less than 2 years (See Appendices). Additionally, the presented analysis shows that fedratinib arm is associated with a gain in survival of more than 3 months. We therefore consider that fedratinib meets the end-of-life criteria.

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	Company's base case before technical engagement	Change(s) made in response to technical engagement	Impact on the company's base-case ICER
Company's preferred base case following technical engagement	Incremental costs: £6,753 Incremental QALYs: 0.60 years ICER: £11,251	Incremental costs: £21,172 Incremental QALYs: 0.48 years ICER: £44,332	Many of the model changes made in response to the technical engagement phase are intertwined. It was therefore not possible to isolate the impact of each change on the overall ICER. Instead we have provided the overall change. See table 5, 6 & 7 and Figures 15, 16, 17, 18 19 & 20 in Appendix for detailed breakdown of results.
Key issues relating to changes in ICER:			
Key issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT	The outputs of the ITC were on an additive scale and were applied in the model as an additive value used fedratinib as the baseline	The outputs of the ITC were in the form of log-odds and were applied as an odds ratio in the model using fedratinib as the baseline	

Key issue 4: Inappropriate approach to modelling	The approach to model TTD and OS events in the original submitted model derived the timing of both events as completely independent of one another.	The changes made in response to technical engagement are detailed in the table above. The new base case fedratinib TTD from the model outcomes is more aligned with the extrapolated TTD than was in the previous base case (See Appendices). The OS in the new base case is calculated from the TTD event as opposed to directly from the OS parametric extrapolation. The base case settings were selected to align with the KM of the selected data sources (See Appendices). The AML health state has been removed.
Key issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	Patients in the BAT treatment arm received BAT until death. Patients in the fedratinib arm received post-fedratinib BAT until death. Patients entering BAT after fedratinib could experience a second response assessment and related utility increments.	The updated model has included a supportive care health state that uses the same resource use assumptions as were applied in the TA386 submission. The HRQoL assumptions for supportive care were also taken from TA386 where a decrement was every 24 weeks until AML or death. Data from HMRN (2020) was used to estimate TTD for the BAT arm and modelling the time in the supportive care state between the points of BAT TTD and OS. Because of the multiple post-fedratinib transitions to BAT, supportive care and death, the movement between these transitions were informed by clinical opinion.

References

1. Harrison CN, Schaap N, Vannucchi AM, et al. Fedratinib in patients with myelofibrosis previously treated with ruxolitinib: An updated analysis of the JAKARTA2 study using stringent criteria for ruxolitinib failure. *American Journal of Hematology*. 2020.
2. Harrison CN, Vannucchi AM, Platzbecker U, et al. Momelotinib versus best available therapy in patients with myelofibrosis previously treated with ruxolitinib (SIMPLIFY 2): a randomised, open-label, phase 3 trial. *Lancet Haematol*. 2018; 5(2):e73-e81.
3. Palandri F, Palumbo GA, Bonifacio M, et al. Baseline factors associated with response to ruxolitinib: an independent study on 408 patients with myelofibrosis. *Oncotarget*. 2017; 8(45):79073-86.
4. Vannucchi AM, Kantarjian HM, Kiladjian JJ, et al. A pooled analysis of overall survival in COMFORT-I and COMFORT-II, 2 randomized phase III trials of ruxolitinib for the treatment of myelofibrosis. *Haematologica*. 2015; 100(9):1139-45.
5. Verstovsek S, Kantarjian HM, Estrov Z, et al. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. *Blood*. 2012; 120(6):1202-9.
6. Gupta V, Egyed M, Hus M, et al. Momelotinib Dose-Intensity is Maintained in JAKi Naive and Previously JAKi Treated Intermediate/High-Risk Myelofibrisis Patients. European Hematology Association. 2020. EP1103.
7. Miller CB, Komrokji RS, Mesa RA, et al. Practical measures of clinical benefit with ruxolitinib therapy: an exploratory analysis of COMFORT-I. *Clinical Lymphoma Myeloma and Leukemia*. 2017; 17(8):479-87.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

APPENDIX

Additional evidence

Key Issue 2

For the technical engagement (TE) economic model, relative treatment effects were applied as odds ratios. The same data and prognostic variables were used as the original submission; however, the outputs and how they were applied were different. As was in the original submission, the base case was informed by the ITC for the 'spleen or symptom' response. Because the number of 'spleen or symptom' responders was not reported in the available PERSIST-2 or SIMPLIFY-2 data, a minimum and maximum analysis of the possible values was run. In the original economic model, a mean of the minimum and maximum analyses was calculated to produce the base case value. However, for the odds scale and to better represent uncertainty, the mean of the maximum and minimum values was estimated before performing the statistical analysis, rather than within the Excel model itself, which allowed confidence intervals to be outputted.

The results of the base case 'spleen or symptom' mean analysis for the intention-to-treat (ITT) population is presented in Table 1. Other analyses, such as for alternative definitions of response for the ITT and cohort 1a populations were repeated and are included as options within the model.

In the base case, the odds ratio was applied to a baseline of 49.38% 'spleen or symptom' response for the fedratinib arm (JAKARTA-2, intermediate-2/high-risk) to produce the proportion of patients responding in the BAT arm. There was consistency in the proportion of patients responding across the ITC approaches included. The 'MAIC - SIMPLIFY-2 (with ECOG PS, DIPSS)' analysis was used as the base case. The calculated response values were not dissimilar to the BAT response values used in the original submission.

Table 1: ITC adjustment for BAT - Odds ratio

ITC approach	ITC adjustment for BAT - Odds ratio (CI)	Calculated BAT response in the model
Naïve - PERSIST-2	7.7 (2.4 – 24.3)	11.26%
Naïve - SIMPLIFY-2	8.8 (3.1 – 24.9)	10.02%
MAIC - SIMPLIFY-2 (with ECOG PS, DIPSS)	8.7 (5.6 – 13.7)	10.04%
STC - SIMPLIFY-2 (with ECOG PS, DIPSS)	8.8 (3.1 – 25.5)	9.95%
MAIC - SIMPLIFY-2 (with ECOG PS, DIPSS, transfusion dep)	9.9 (2.7 – 20.2)	8.99%

Key Issue 3

Data from the ongoing global chart review was used as an additional information source to inform overall survival from the point of resistance/refractory/intolerance (r/r/i) or progression

The available survival data comprised intermediate-2 and high-risk patients who had experienced r/r/i to ruxolitinib and /or progression after ruxolitinib initiation and had either remained on ruxolitinib or discontinued. OS data was provided from the point of progression if it was available, or r/r/i if it was not. The outcomes of [redacted] patients were reported, [redacted] patients had a recorded progression event and [redacted] patients experienced a r/r/i event. Because the baseline for the time to OS event was taken from the point of progression if the data was available, only [redacted] of the patients had their baseline taken from the point of r/r/i.

The data suggested that ruxolitinib did not extend overall survival for patients who continued rather than discontinued ruxolitinib at progression or r/r/i (Figure 1). For patients with their OS baseline provided from the point of r/r/i, this relationship held true, albeit with a smaller sample size (Figure 2).

Two scenarios to inform OS were therefore included based on this data:

1. The patients who continued ruxolitinib in Figure 1 were used to produce parametric models that were included as options to inform BAT OS events (option for responders and non-responders)

2. All patients from Figure 2 were used to produced parametric models that were included as options to inform BAT OS events (option for responders and non-responders) – All patients was selected due to the limited number at risk and the lack of difference between the patients who discontinued ruxolitinib and those who did not.

Schain et al. remained as the base case option informing OS in the model. The additional parametric extrapolations used as modelling options are presented in Figure 3 and Figure 4.

[a] Patients were considered to have progressed as defined and documented by the treating clinician.
[b] Patients were considered to have relapsed if after ruxolitinib initiation and after 3 months of treatment with ruxolitinib, spleen size decreased and then subsequently increased (e.g., mild splenomegaly [at ruxolitinib initiation] to no splenomegaly to moderate splenomegaly); patients were considered to have refractory disease if after ruxolitinib initiation and after 3 months of treatment with ruxolitinib, spleen size increased without documentation of spleen size reduction (e.g., mild splenomegaly [at ruxolitinib initiation] to moderate splenomegaly); patients who continued to have severe splenomegaly throughout ruxolitinib treatment were considered to have refractory disease if treating physicians indicated the patient to have progressed while on treatment with ruxolitinib.
[c] Patients were defined intolerant to ruxolitinib if they were on ruxolitinib for ≥28 days, discontinued ruxolitinib within 90 days, and had an adverse event resulting in anemia, thrombocytopenia, or had a progressive disease with regard to anemia (including transformation to acute myeloid leukemia [AML]) as the reason for ruxolitinib discontinuation.

Figure 1: OS from point of progression or r/r/i of patients

Figure 2: Chart review OS from patients where baseline == r/r/i

Figure 3: Chart review parametric extrapolations – all patients who continued ruxolitinib after r/r/i or progression

Figure 4: Chart review parametric extrapolations – all patients with r/r/i as OS baseline

Key Issue 4

In response to the ERG report, structural changes were made to the model that assume dependency between TTD and OS events. Firstly, TTD is now estimated from model entry, rather than using inputs split pre- and post- 24 weeks. For OS, two options have been included that introduce dependency between OS and TTD. One approach estimates fedratinib OS from the point of discontinuation, and the second approach uses a common random number to derive both TTD and OS events. In addition, the OS and TTD outcomes for both the fedratinib arm and the BAT arm are modelled according to the 24-week response outcome. An updated model diagram showing the changes to the model structure is presented in Figure 5.

The data for TTD and OS from JAKARTA-2 intermediate-2/high risk patients is the same data that was used to inform the original pre- and post- 24 weeks parametric models, but analysed to produce results from a baseline of week 0. Additionally, JAKARTA-2 data has been used to model OS from the point of TTD.

The base case is to model TTD from week 0 split by responders and non-responders (Figure 6 and Figure 7); OS for fedratinib is then modelled from the point of discontinuation split by responders and non-responders (Figure 8 and Figure 9). The final TTD curves derived in the model are presented in Figure 10 and show the similarity to the selected parametric extrapolation. The base case response

definition was ‘spleen or symptom’ response; the equivalent parametric extrapolations for the full population with either ‘spleen’ or ‘symptom’ response have also been included as options in the model.

Figure 5: Updated model structure

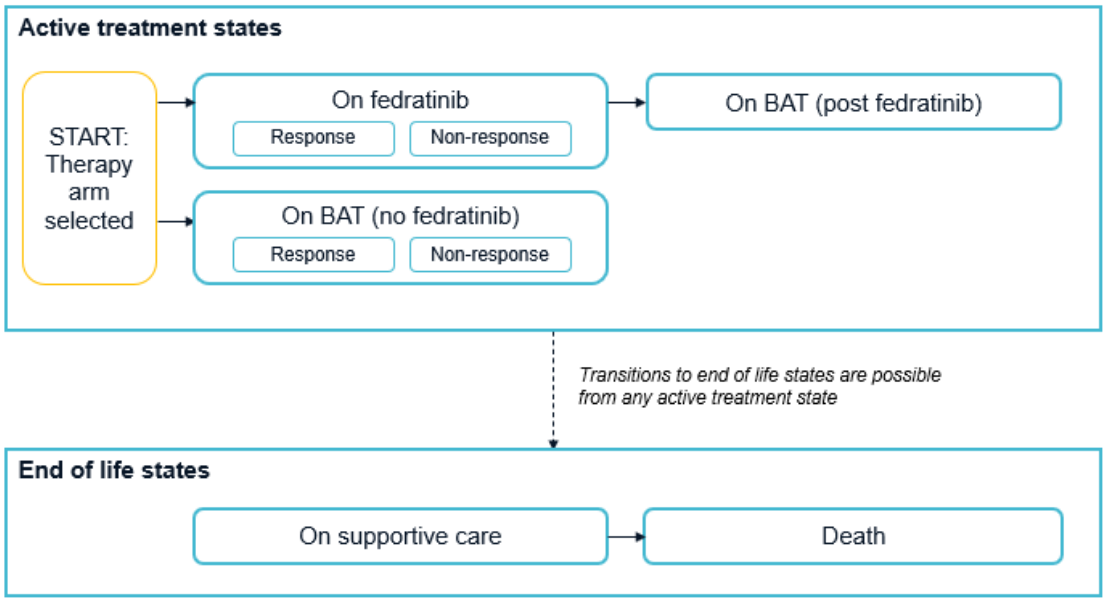


Figure 6: JAKARTA-2 TTD from 0 weeks – responders

Figure 7: JAKARTA-2 TTD from 0 weeks – non-responders

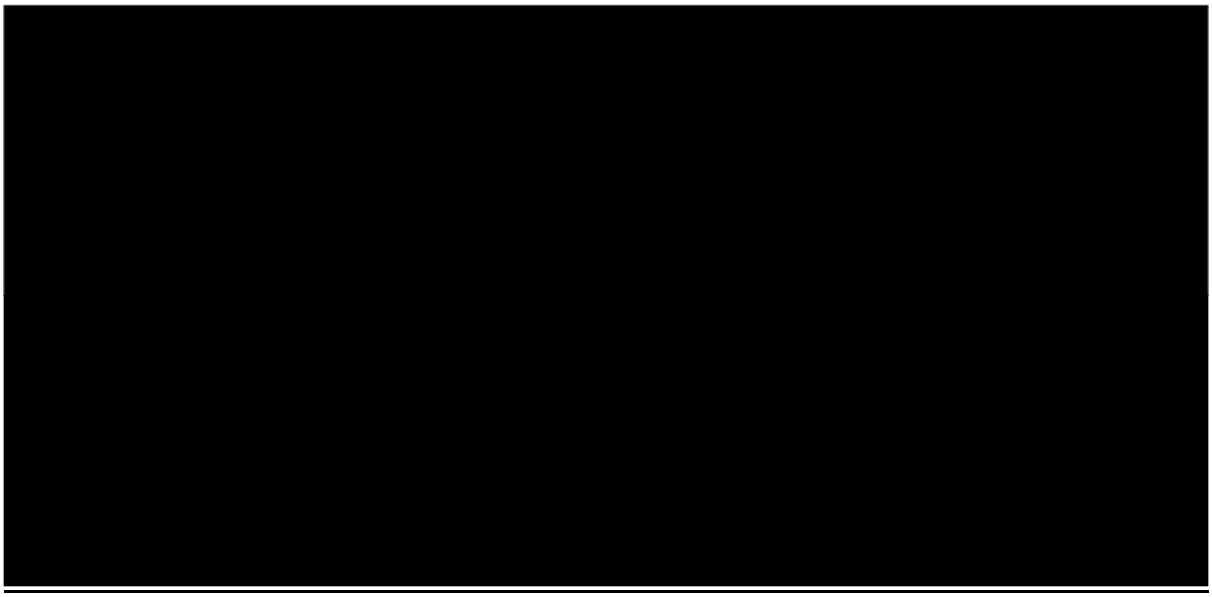


Figure 8: JAKARTA-2 OS from TTD event – responders

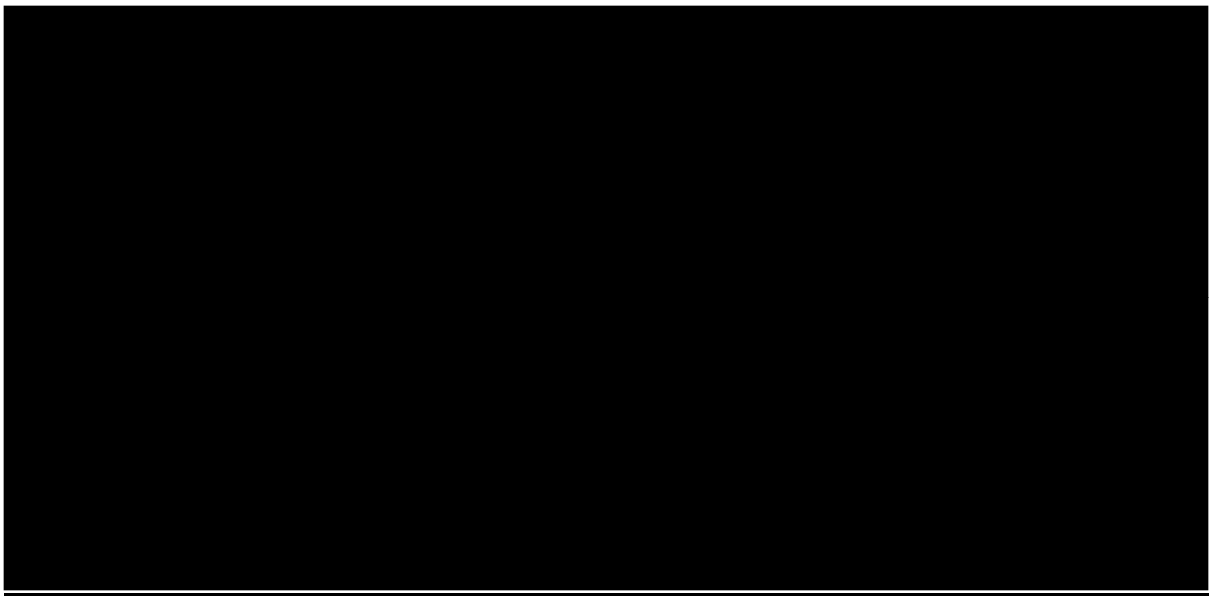


Figure 9: JAKARTA-2 OS from TTD event – non-responders

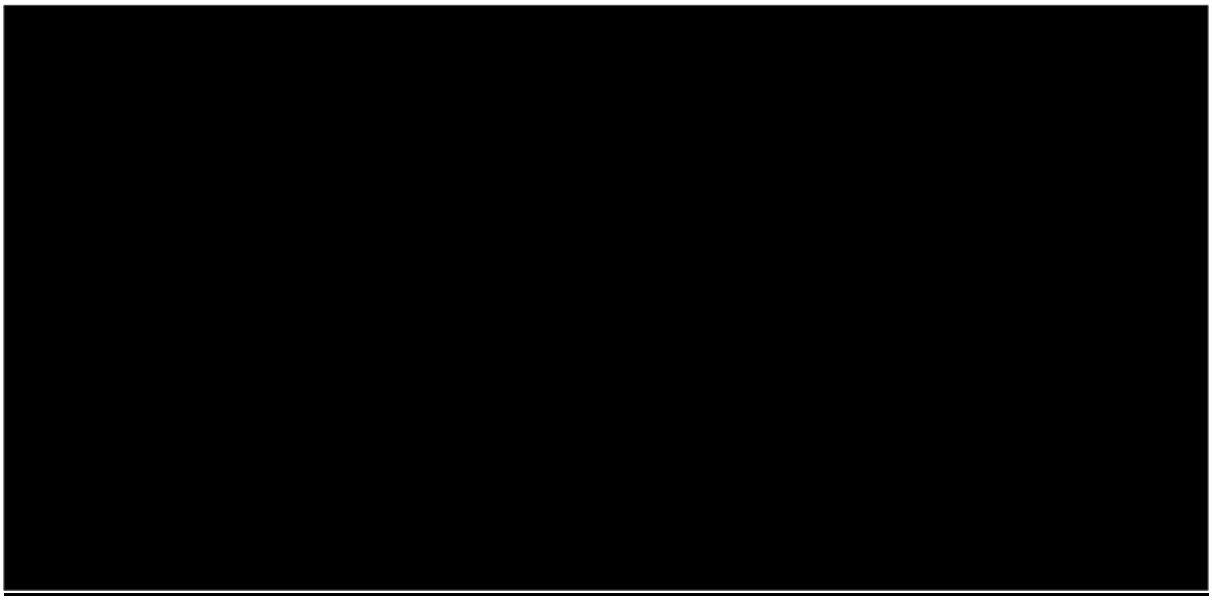


Figure 10: JAKARTA-2 modelled base case TTD

Key Issue 5

The model was updated to include a supportive care health state that uses the same resource use assumptions as were applied in TA386. The HRQoL assumptions for supportive care were also taken from TA386.

The additional evidence includes the weekly resource use for supportive care, which is presented in Table 2; which takes the base case resource use for BAT that was originally in the model and applies the assumptions from TA386. The costs per resource are equivalent to the original submitted model. An ongoing 24-weekly utility decrement for patients in supportive care of -0.025 (SMC submission for ruxolitinib, given the value is redacted in TA386) was applied equally between patient arms.

New data from HMRN (2020) was provided to estimate TTD for the BAT arm. Post-fedratinib transitions to BAT, supportive care and death were informed by inputs from clinical opinion.

The HMRN provided pooled second-line TTD data for patients with myelofibrosis, which was used to fit parametric extrapolations for the base case BAT TTD. A limitation of these data is that the TTD is not necessarily in relation to ruxolitinib, but to any treatment received as second-line for myelofibrosis. These extrapolations are presented in Figure 11.

Clinical opinion for post-fedratinib transitions were that for non-responders, [redacted] would be expected to continue to BAT post-fedratinib, with [redacted] transitioning to supportive care. For responders, it was estimated that [redacted] patients would transition to BAT post-fedratinib, with [redacted] transitioning to supportive care.

For those patients who receive BAT post-fedratinib, the proportion of remaining time alive spent in supportive care versus BAT was estimated to be [redacted] for both responders and non-responders, based on the ratio of undiscounted life years between BAT and supportive care observed in the BAT arm results.

Table 2: Supportive care resource use applied in model

Resource	Use per week	Source
A&E visit	0.013	Assumed equal to BAT, TA386
FBC & U&E	0.160	50% lower than BAT, TA386
Hospital night	0.150	Assumed equal to BAT, TA386
Outpatient visit	0.110	50% lower than BAT, TA386
Primary care visit	0.030	Assumed equal to BAT, TA386
RBC unit transfusion	0.190	COMFORT-I placebo arm, TA386
Urgent care	0.003	Assumed equal to BAT, TA386

Figure 11: The HMRN pooled 2L TTD parametric extrapolations

Key Issue 6

The inconsistency in assumptions raised by the ERG between fedratinib and BAT have been addressed in Key Issues 4 and 5. All additional data contributing to the improvement of consistency has been outlined in the sections above.

Key Issue 7

OS data for BAT in SIMPLIFY-2 was presented in a slide deck developed by Sierra Oncology. As OS was not a specified endpoint in the SIMPLIFY-2 trial, this was a post-hoc analysis. It was assumed initially that the data would be reported correctly by the authors. However, due to discrepancies within the source itself, further assessment was conducted, and the accuracy of this evidence is uncertain given the following:

- The slide deck was found through a Google search and the OS evidence has not been presented in an accredited, peer-reviewed journal
- The text describing OS in SIMPLIFY-2 refers to a double-blind treatment phase but SIMPLIFY-2 was open-label, casting doubt on the accuracy of the data reported

- Assuming the author of the slide deck meant 24 weeks when referring to the double-blind treatment phase, the text describing OS at the end of 24 weeks does not match the Kaplan-Meier presented:

The text stated that 21% of BAT subjects had died after 24 weeks, whereas the pseudo Kaplan-Meier data indicates that 4 of the 52 patients had died during the first 24 weeks (8%)

Further, as patients in the BAT arm of SIMPLIFY-2 could receive momelotinib after 24 weeks, the data after 24 weeks is not representative of patients receiving BAT. The pseudo patient-level data was therefore censored at the end of Week 24. 24 weeks is not considered a long enough time to assess overall survival data.

While the analysis is not considered appropriate for the above reasons, it is provided below at the request of the ERG. Matching was performed firstly on DIPSS, and then on as many of the DIPSS Plus items as the availability of data allowed:

- Age (mean and standard deviation [SD])
- Haemoglobin (mean and SD)
- Transfusion dependence
- Platelet count (Mean SD)

After matching, average baseline characteristics were balanced across the JAKARTA-2 and SIMPLIFY-2 patients but the effective sample size (ESS) was relatively low (32.1). Given the large difference in the numbers of patients who are transfusion dependence across the two studies, there are a small number of transfusion dependent patients in the JAKARTA-2 study with high weights contributing heavily to the final analysis.

Table 3: Exploratory OS MAICs with SIMPLIFY-2 BAT arm in first 24 weeks

Method	HR (95% CI)	JAKARTA-2 N / ESS
Naïve	[redacted]	[redacted]
MAIC (matching on DIPSS)	[redacted]	[redacted]
MAIC (DIPSS: matching on age and haemoglobin)	[redacted]	[redacted]
MAIC (DIPSS plus: matching on age, haemoglobin, transfusion dependence and platelet counts)	[redacted]	[redacted]

Figure 12: OS MAIC with SIMPLIFY-2 BAT arm in first 24 weeks (matching on DIPSS)

Figure 13: OS MAIC with SIMPLIFY-2 BAT arm in first 24 weeks (matching on age and haemoglobin)

Figure 14: OS MAIC with SIMPLIFY-2 BAT arm in first 24 weeks (matching on age, haemoglobin, transfusion dependence and platelet counts)

Key Issue 8

The issue that the ERG raised on the lack of face validity of the stopping rule scenario has been addressed with the addition of the new model structure where the OS for fedratinib is dependent on the time spent on fedratinib. The additional data is provided in Key Issue 4.

Key Issue 9

The mean platelet count for both SIMPLIFY-2 arms was lower than the JAKARTA-2 mean platelet count, so the ERG were concerned that the costing of ruxolitinib would not be representative of the UK MF population. The SIMPLIFY-2 study did not publish the proportion of patients with a platelet count below 100,000 per µl but did publish the mean platelet count and standard deviation (provided in Table 4). A normal distribution was assumed. As such, the NORM.DIST() function in Excel was populated with the mean and standard deviation to calculate the proportion of patients having a platelet count below the 100,000 per µl threshold.

The platelet count distributions for both SIMPLIFY-2 treatment arms were included as scenarios in the model for calculating the proportion of patients below the 100,000 per µl threshold. Both values are broadly similar to the 34.6% value from JAKARTA-2 used in the model base case.

Table 4: Informing the % Estimated patients with <100,000 platelets/µl with SIMPLIFY-2 data

SIMPLIFY-2 treatment arm	Mean platelet count	SD	% Estimated <100,000 per µl
BAT	126,500 per µl	95,900	39.1%
Momelotinib	170,800 per µl	148,000	31.6%

Key Issue 10

Relative treatment effects have been updated in the model to be odds ratios. The additional evidence for this has been described in Key Issue 2

Key Issue 11

The calculation of OS for BAT and fedratinib in the TE response model is described in Key Issue 4. No additional evidence has been added to the model.

Results

Base-case incremental cost-effectiveness analysis results

Table 5: Base-case results (based on net price)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	2.844	1.430	-	-	-	-	-
Fedratinib	[redacted]	3.595	1.908	21,172	0.752	0.478	44,332	44,332

Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.

Table 6: Disaggregated outcomes

	Treatment arm		Increment	Absolute increment	% absolute increment
	BAT	Fedratinib			
Costs by health state (£)					
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Death (End of life)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	100%
Costs by category (£)					
Acquisition	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

	Treatment arm		Increment	Absolute increment	% absolute increment
	BAT	Fedratinib			
Administration	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adverse events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Resource use	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Thiamine testing and supplementation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
End of life	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	100%
Life years (LYs)					
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	100%
Median	[redacted]	[redacted]	[redacted]	-	-
Quality-adjusted life years (QALYs)					
JAKi state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BAT state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

	Treatment arm		Increment	Absolute increment	% absolute increment
	BAT	Fedratinib			
Supportive care state	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	100%
Median	[redacted]	[redacted]	[redacted]	-	-
Key: Acute myeloid leukaemia; BAT, best available therapy; JAK, Janus kinase; JAKi, JAK inhibitor; QALY, quality-adjusted life year.					

Probabilistic sensitivity analysis

Table 7: Probabilistic sensitivity analysis results (based on net price)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER vs baseline (£/QALY)	ICER incremental (£/QALY)
BAT	[redacted]	3.282	1.478	-	-	-	-	-
Fedratinib	[redacted]	5.212	2.224	30,959	1.929	0.746	41,520	41,520
Key: BAT, best available therapy; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.								

In probabilistic analysis, the LYG on both treatment arms is higher. For fedratinib, this is a result of the skewed and uncertain survival distribution, namely for responders. For BAT, this is a result of the assumption that BAT survival follows fedratinib survival in the long-term, to avoid the curves crossing. Overall, the ICER is similar between deterministic and probabilistic analyses.

Figure 15: Cost-effectiveness plane – Fedratinib vs BAT

Figure 16: Cost-effectiveness acceptability curve – Fedratinib vs BAT

Deterministic sensitivity analysis

It was considered important to include parametric curves within the deterministic sensitivity analysis, given that these inputs are key to the cost-effectiveness of fedratinib. Curve parameters were varied jointly in deterministic sensitivity analysis by assuming the random numbers for all parameters in the multivariate normal sampling to be at 0.025 for the lower bound and 0.975 for the upper bound. This is a pragmatic approach with the intention of showing the impact of parameter uncertainty for these curves.

As a result of the uncertainty in the fedratinib responder OS curve (post-discontinuation), the lower-bound yields a negative ICER (due to the reduction in LYs and therefore QALYs). Therefore, incremental net monetary benefit (INMB) is also presented to better interpret deterministic sensitivity analysis results in the presence of negative ICERs.

Figure 17: Results of one-way sensitivity analysis (ICER)

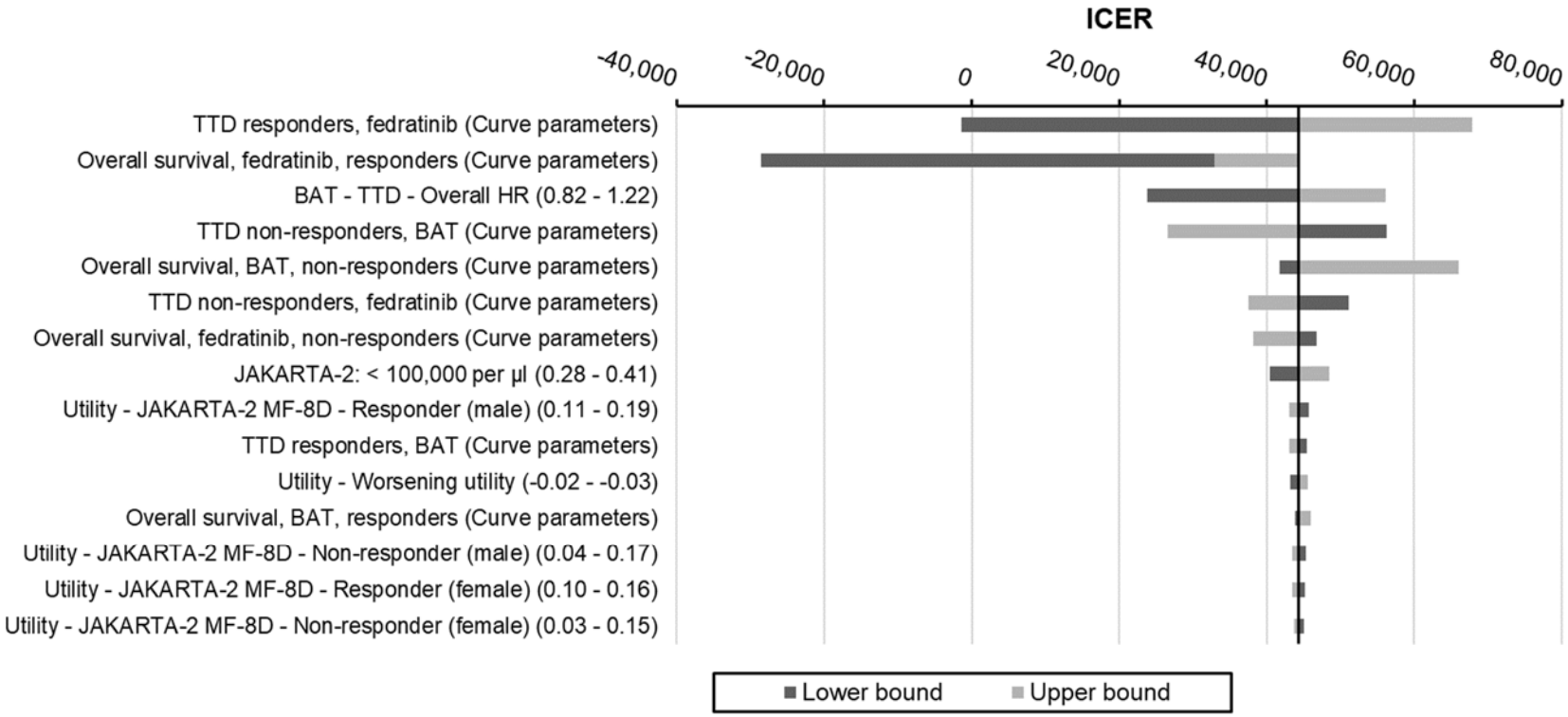
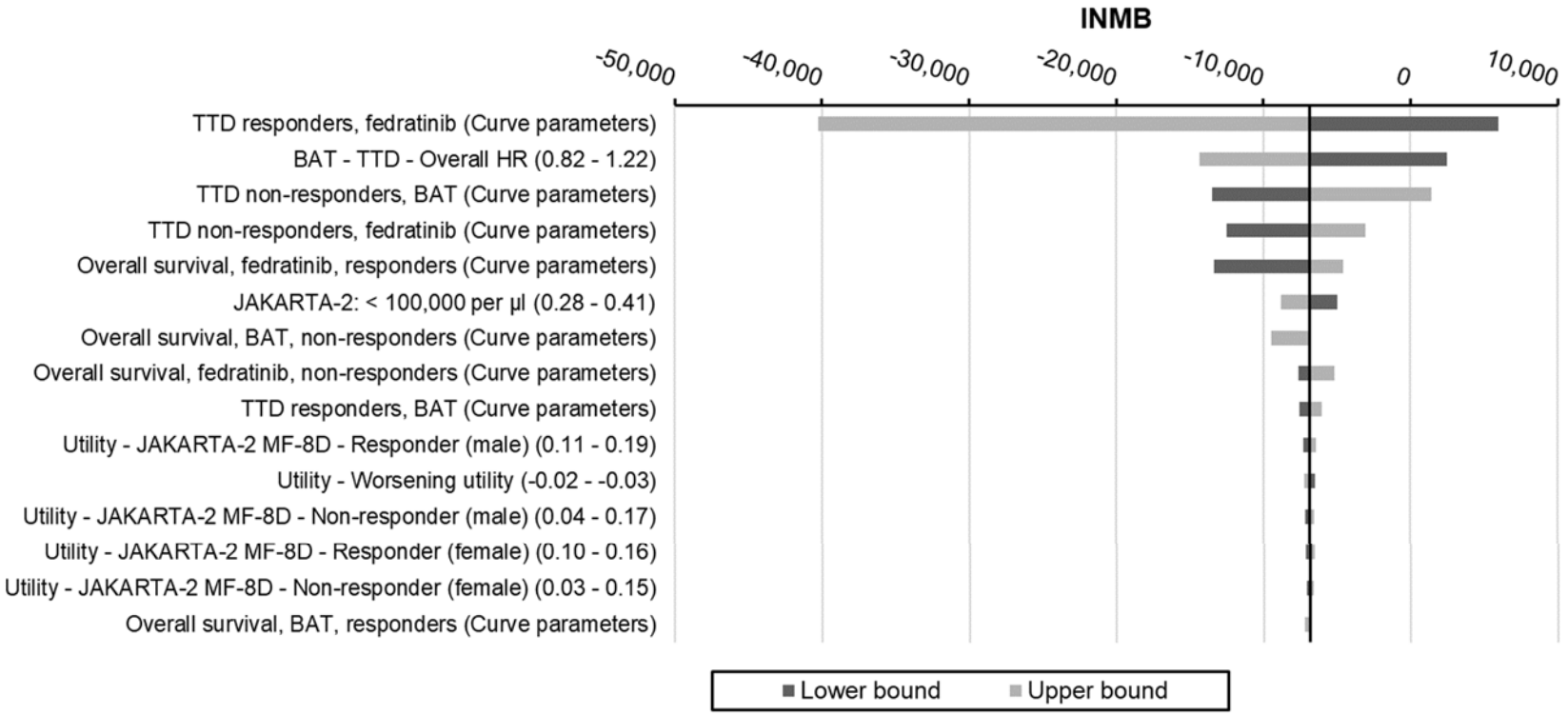


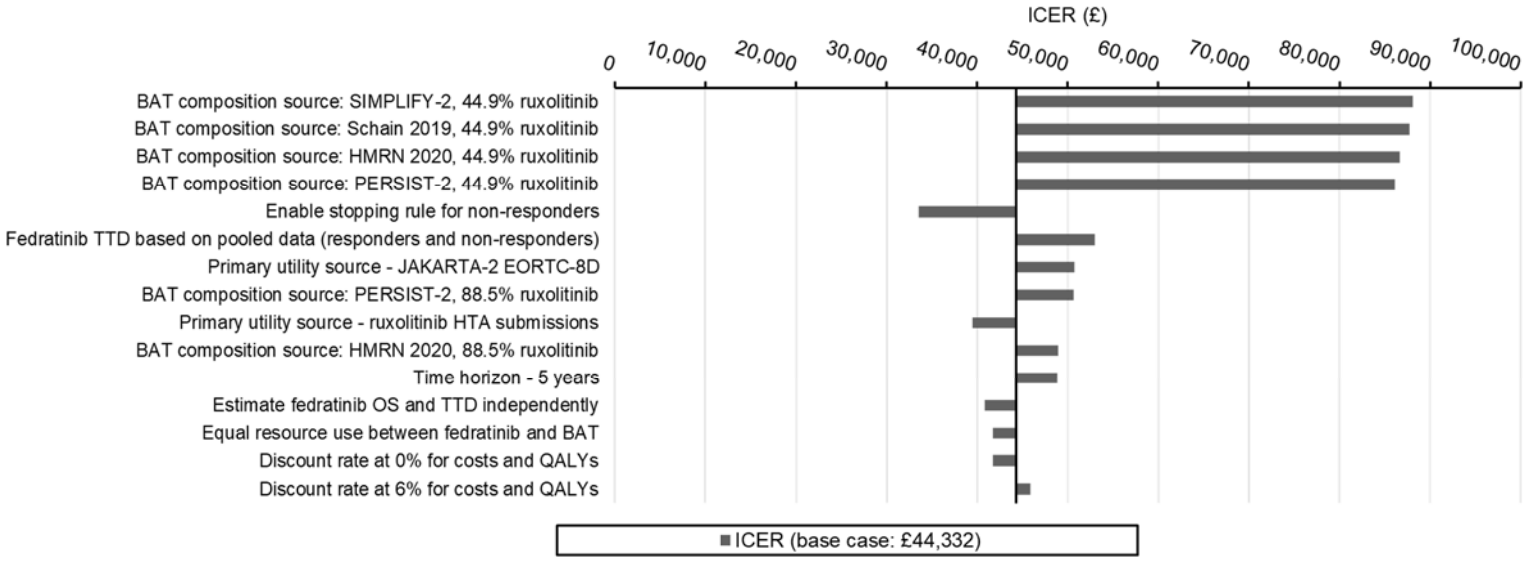
Figure 18: Results of one-way sensitivity analysis (INMB)



A willingness-to-pay threshold of £30,000 per QALY is assumed for the INMB diagram.

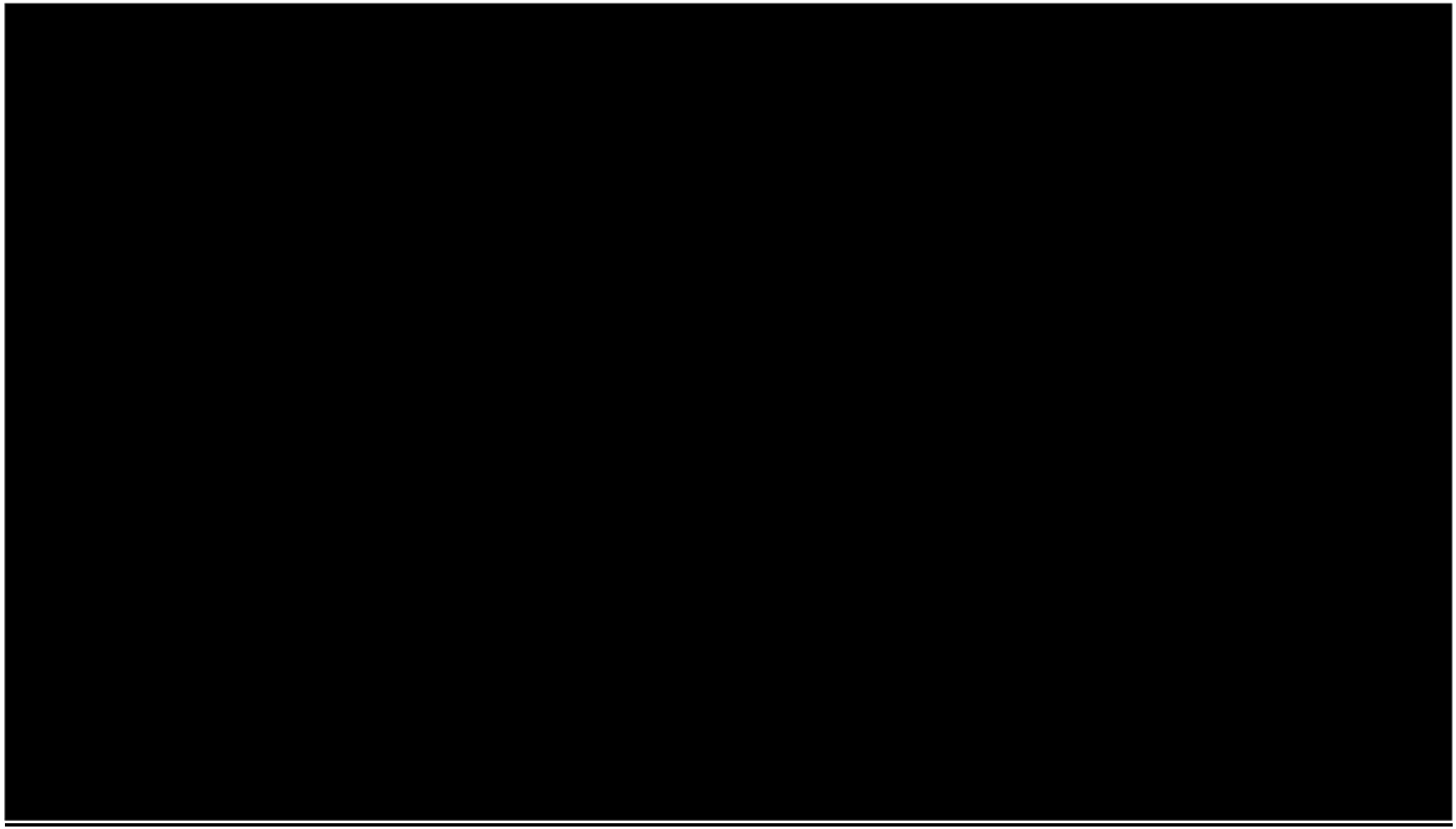
Scenario analysis

Figure 19: Summary of modelling scenarios which had the most impact on the base case ICER



Base-case clinical results

Figure 20: Modelled fedratinib OS versus KM data



NICE National Institute for Health and Care Excellence

Clinical expert statement & technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

Thank you for agreeing to comment on the ERG report for this appraisal, and for providing your views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature. The ERG report and stakeholder responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

Information on completing this form:

- In **part 1** we are asking you to complete questions where we ask for your views on this technology. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.
- In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.
- The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a clinical perspective could help either:
- resolve any uncertainty that has been identified
  OR
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

Please return this form by **5pm** on **Thursday 5 November 2020**

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Important information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow.If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

PART 1 – Treating a patient with myelofibrosis and current treatment options	
About you	
1. Your name	**Professor Claire N Harrison**
2. Name of organisation	**Guy's and St Thomas' NHS Foundation Trust**
3. Job title or position	**Consultant Haematologist**
4. Are you (please tick all that apply):	☐ an employee or representative of a healthcare professional organisation that represents clinicians? x a specialist in the treatment of people with this condition? ☐ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	**x** yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☐ other (they didn't submit one, I don't know if they submitted one etc.)

6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐ yes **No I did not write the organisation submission**
7. Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**none**
The aim of treatment for this condition	
8. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	To reduce the spleen size as a surrogate of disease activity and secondly to improve symptoms and quality of life.
9. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm,	Reduction of spleen volume by at least 10% (previously shown in studies with Ruxolitinib to be lined with survival benefit). The endpoint in clinical trials is 35% but an ad hoc analysis of the COMFORT trial data suggests that at least a 10% reduction is a surrogate of survival benefit. Even in the absence if spleen volume reduction improvement in disease related symptoms and quality of life is an important endpoint.

or a reduction in disease activity by a certain amount.)	
10. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Yes. The previous studies demonstrated in the first line setting that best available therapy was inferior to ruxoltinib. In the second line setting there is no rationale for believing that best available therapy would work. No other approved therapies exist. Very few patients would be eligible for stem cell transplant and usually they would receive this before they lose their response to ruxolitinib.
What is the expected place of the technology in current practice?	
11. How is the condition currently treated in the NHS?	Either by continuing Ruxolitinib or switching to a range of best available therapy. Often clinicians do not discontinue Ruxolitinib as patients often get worsened symptoms and spleen enlargement (so-called Ruxolitinib withdrawal syndrome if the drug is withdrawn).
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	Yes ELN, BSH and NCCN guidelines. Only NCCN guidelines will reflect Fedratinib
• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	The pathway of care is quite well defined, however the physician population is not used to defining ruxolitinib intolerance or resistance, dosing is an important element particularly under-dosing of ruxolitinib leading to suboptimal benefit.
• What impact would the technology have on the current pathway of care?	Patients will switch from one tablet to another, I do not believe this will alter the current pathway of care.

12. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Yes
• How does healthcare resource use differ between the technology and current care?	As the patients physical condition improves they require less hospital intervention. In the absence of the technology the patients with either continue suboptimal Ruxoltiinib and progress or come off and be on supportive care gradually progressing. This is what happened to my patients on the JAKARTA2 trial after the withdrawal of Fedratinib. There is indirect evidence that Fedratinib treated patients require less hospital based care.
• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Specialist clinics in secondary care.
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	None
13. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes for the duration of response.

• Do you expect the technology to increase length of life more than current care?	Yes as the technology reduces spleen size which has been linked to improvement of survival unfortunately the design of the clinical trials as affected by the full clinical hold for this agent means that prolongation of survival was no demonstrated.
• Do you expect the technology to increase health-related quality of life more than current care?	Yes evidence from the clinical trials and direct observation of my own patients suggests that this would be true.
14. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	No
The use of the technology	
15. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed, additional clinical requirements, factors	The technology is quite straightforward to use. It will be more simple to use than treatments requiring for example transfusion or monitoring for example for side effects of thalidomide. It does however require monitoring of thiamine levels and management of gastrointestinal toxicity which in my experience is simple to do and for the latter only required early in the treatment course.

affecting patient acceptability or ease of use or additional tests or monitoring needed.)	
16. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	No particular rules, monitoring of spleen or symptoms is routinely performed. The patient will need to be reviewed for thiamine deficiency which is specific to this medication.
17. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	No I think the benefits will be included in the QALY
18. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it	Yes there is currently no approved therapy to be used after failure of Ruxolitinib.

improve the way that current need is met?	
• Is the technology a 'step-change' in the management of the condition?	Yes nothing currently available with exception of agents already proven to be of no benefit.
• Does the use of the technology address any particular unmet need of the patient population?	Yes there is no current alternative treatment in the second line setting which is therefore an area of major unmet need.
19. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	Yes potentially Wernicke's encephalopathy would be life threatening or chaning. Fedratinib like Ruxolitinib causes anaemia and in addition it can cause some GI side effects which in my experience is easily managed.
Sources of evidence	
20. Do the clinical trials on the technology reflect current UK clinical practice?	Yes however the trials had to be prematurely stopped
• If not, how could the results be extrapolated to the UK setting?	NA

• What, in your view, are the most important outcomes, and were they measured in the trials?	Safety, Spleen and symptom response. These were measured in the trials
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	Yes these are the accepted outcome measures in this clinical setting. Unfortunately long term efficacy and safety data is not available from the JAKARTA study. Data is however available from the earlier phase I and II clinical trials.
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	Yes risk of Wernicke's Encephalopathy which was evaluated. There is a black box warning for this.
21. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	Yes I have personal experience of treating patients in the second line setting with fedratinib and what happened to them after the JAKARTA 2 study was stopped. All of these patients have now died apart from one who is undergoing treatment for accelerated phase disease. The remaining patients had all died between 2-3 years of starting on the JAKARTA2 trial.
22. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance 386?	Yes from the SIMPLIFY 2 study where the efficacy of other therapies after Ruxolitinib or indeed continuing Ruxolitinib were compared to Momelotinib. In addition the PERSIST2 trial.

23. How do data on real-world experience compare with the trial data?	NA
Equality	
24a. Are there any potential equality issues that should be taken into account when considering this treatment?	No major equality issues.
24b. Consider whether these issues are different from issues with current care and why.	Not relevant
Topic-specific questions	
25. For people with myelofibrosis who are relapsed/refractory to ruxolitinib, what treatments are used as part of BAT?	Staying on Ruxolitinib, Hydroxycarbamide, Interferon, ESA, Thalidomide, Steroid, Danazol, supportive care. Rarely these patients might move to stem cell transplant
26. In clinical practice, do people continue to have BAT for a lifetime? If BAT is stopped, when	BAT or on-going Ruxolitinib would be used generally until the time of patient death

in the treatment pathway does this usually happen, and what treatments are offered next?	
27. Are the following groups considered to be clinically distinct populations in current NHS practice? (a) People with disease that is relapsed/refractory to ruxolitinib. (b) People who stop ruxolitinib because they cannot tolerate it or have adverse events.	Current NHS practice does not include a string definition of relapse or refractory myelofibrosis in the context of Ruxolitinib, data suggests these patients have a poor prognosis. Patients who stop Ruxolitinib or cannot tolerate it are likely to represent a totally different group.
28. In current practice, for each of the above populations, how many people would continue to have ruxolitinib, rather than stopping it and switching to something else?	a) most patients >80% would continue b) most >90% would discontinue
29. Similarly, would you expect treatment with fedratinib to	I think that if patients were on fedratinib and their disease worsened or relapsed they may stay on the drug but I suspect most would switch back to Ruxolitinib or a clinical trial.

continue after relapse, rather than stopping it and switching to something else?	
30. If someone becomes refractory to ruxolitinib but continues to have it, would you expect them to have better outcomes compared with if they did not continue ruxolitinib?	This would be dependent on what the patient had for treatment after ruxoliitinib. Stopping Ruxolitinib suddenly is life threatening I would expect those completely discontinuing without moving on to fedratinib or another therapy would result in significantly worse outcomes.
31. Are one or both of spleen length and spleen volume used to measure response?	To measure response reliably spleen volume is required, spleen length is highly variable.,
32. Thinking about current NHS practice: what is the average life expectancy for people with myelofibrosis who are relapsed/refractory to ruxolitinib? What is the expected survival proportion at 5 years and 10 years?	<2 years… 7/8 of the JAKARTA 2 patients had died within 2 -3 years after discontinuing the therapy even though the majority went back to ruxoltiinib. By 5 years I would expect this to be over 90% I would not expect any patients to be alive at 10 years.

33. How generalisable is the population in the JAKARTA-2 (fedratinib) trial to people with myelofibrosis who are relapsed/refractory to ruxolitinib treated in typical NHS practice? How would you expect the trial population's life expectancy to compare with that of the population you see in current NHS practice (i.e. your answer to 32)?	Since the eligibility criteria for JAKARTA2 were quite unrestricting I think the patient population is generalizable I think the life expectancy would be similar to the current NHS population.

PART 2 – Technical engagement questions for clinical experts	
Issues arising from technical engagement	
We welcome your response to the questions below, but you do not have to answer every question. If you think an issue that is important to clinicians or patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the professional organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.	
Key issue 1: Concerns with phase 2, single-arm JAKARTA-2 study	Clinical hold limits longer-term response duration. Criteria for Ruxolitinib intolerance or resistance are yet to be fully established. The trial data has been reanalysed with more recent criteria.
Key issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT	This is a limitation of the data nonetheless there are other studies in this setting

Key issue 3: Alignment between the comparator and the modelled population	I agree with the modelled states.
Key issue 4: Inappropriate approach to modelling	Very difficult to have a straightforward approach to modelling.
Key issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	Very few patients are on no therapy even in the end of life, palliative phase. This is because the disease is linked to very bad quality of life and clinicians will frequently continue therapies to reduce suffering.
Key issue 6: Inconsistent assumption between BAT and fedratinib	No comment
Key issue 7: Assumption of survival difference	The model for survival benefit in my opinion is reasonable. There is good concordance between spleen response and to lesser extent symptom response and survival benefit. Given that these are incrementally improved with Fedratinib this would seem reasonable.

Key issue 8: Lack of face validity for the stopping rule scenario	It would be extremely difficult indeed in my opinion impossible to validate this.
Key issue 9: Costs for the comparator arm (ruxolitinib)	Most patients will stay on Ruxolitinib as identified in the SIMPLIFY2 study.
Key issue 10: Reliability of response rate	The response rate and of course duration is limited by the full clinical hold. The data has been analysed using more strict criteria for Ruxolitinib intolerance or refractory disease and remains consistent.
Key issue 11: End of life criteria	In my opinion the indication for use after failure or disease refractory to Ruxolitinib meets criteria for End of Life. All bar one of patients whom I treated in the JAKARTA2 trial have died in under 3 years.
Are there any important issues that have been missed in ERG report?	No

PART 3 -Key messages

16. In up to 5 sentences, please summarise the key messages of your statement:

- Patients with myelofibrosis have a disease characterised by a poor prognosis, splenomegaly and disease related symptoms which significantly affect quality of life.

- Ruxolitinib as previously assessed by NICE (HTA386) significantly improves the lives of a majority of these patients but the duration of effect is relatively short c3years, after which time the prognosis of patients is poor, benefits of ruxolitinib correlate with spleen response.
- For patients who have failed ruxolitinib Fedratinib as assessed in the JAKARTA2 study (and indeed first line in the JAKARTA study and earlier phase I/II studies) offers a meaningful option delivering objective reduction in spleen volume and symptom burden.
- Without a second line option patients failing ruxolitinib have no alternative but to either continuing suboptimal ruxolitinib or switch back to what has become known as best available therapy which we know is suboptimal; in second-line clinical trials where staying on ruxolitinib was an option (eg SIMPLIFY 2) ruxolitinib was the treatment of choice.
- Although cross-trial comparisons are fraught with difficulty the response rate to continuing ruxolitinib in the control arm of SIMPLIFY2 was inferior to Fedratinib in JAKARTA2.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed document, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Clinical expert statement & technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

Thank you for agreeing to comment on the ERG report for this appraisal, and for providing your views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature. The ERG report and stakeholder responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

Information on completing this form:

- In **part 1** we are asking you to complete questions where we ask for your views on this technology. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.
- In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.
- The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a clinical perspective could help either:
- resolve any uncertainty that has been identified
  OR
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

Please return this form by **5pm** on **Thursday 5 November 2020**

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Important information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow.If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

PART 1 – Treating a patient with myelofibrosis and current treatment options	
About you	
1. Your name	**Adam Mead**
2. Name of organisation	**University of Oxford**
3. Job title or position	**Professor of Haematology**
4. Are you (please tick all that apply):	☒ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☒ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☒ yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☐ other (they didn't submit one, I don't know if they submitted one etc.)

6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐ yes **N/A**
7. Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**N/A**
The aim of treatment for this condition	
8. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	The aim of the treatment of myelofibrosis varies from patient to patient and is very much personalised. Broadly the treatment aim falls into a number of different areas: 1. To improve disease related systemic symptoms 2. To improve splenomegaly and associated symptoms 3. To improve life expectancy 4. To control myeloproliferation (raised blood counts) 5. To improve cytopenias, particularly anaemia 6. To reduce the risk of disease complications (blood clots, bleeing, infection, transformation to leukaemia etc)

9. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	Tumour is an odd term in this disease context. Easiest to focus response on spleen size. Put simply, the main goal is to make patients feel better and in my view there is no specific goal in terms of spleen size reduction to achieve this aim. Some patients will achieve a marked symptom improvement with minimal improvement in spleen volume. Some studies have suggested that greater spleen volume reductions with JAK2 inhibition correlate with improved long term survival in myelofibrosis i.e. it is an important surrogate marker. Here the aim would be to achieve a 35% reduction in spleen volume approximating to a 50% reduction in palpable spleen length.
10. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Whilst the currently approved JAK2 inhibitor in the NHS (ruxolitinib) is undoubtedly an effective treatment for some patients, there remains a major unmet need for new therapies, particularly for patients who fail to respond adequately to rux or develop side effects. Ruxolitinib is also associated with some important limitations: • Side effects: The main toxicities associated with the use of ruxolitinib are haematological (anaemia and thrombocytopenia) and immune suppression, with a range of different infections reported to be associated with the use of ruxolitinib. For example, in the COMFORT-II study, grade 3 or 4 anaemia occurred in 42% of patients compared with 31% for BAT. All grades of thrombocytopenia occurred in 68% of patients compared with 29% for BAT. We are seeing infections and skin tumour emerge as important side effects, requiring ruxolitinib treatment discontinuation in a number of patients. • Although almost all patients show some reduction in spleen volume with ruxolitinib, the majority of patients with MF fail to achieve ≥ 35% reduction in spleen volume after 24 or 48 weeks of therapy with ruxolitinib. The majority of MF patients continue to have ongoing splenomegaly at the time of best response to ruxolitinib. Many patients have ongoing symptoms despite rux treatment. • A significant proportion of patients lose the response to ruxolitinib with prolonged follow up. The median time to loss of response is approximately 3 years. Patients showing progression on ruxolitinib have a poor prognosis with a major unmet need for new treatments. • Few MF patients show complete reversal of fibrosis, even with prolonged ruxolitinib treatment. • Whilst overall survival of MF patients is somewhat improved with ruxolitinib therapy, molecular (clonal) responses are only infrequently seen with ruxolitinib treatment i.e. the therapy does not induce significant

	disease modification in the majority of patients and disease eradication or "cure" has not been reported with ruxolitinib.
What is the expected place of the technology in current practice?	
11. How is the condition currently treated in the NHS?	The clinical presentation of PMF is highly heterogeneous, with variable splenomegaly, blood count abnormalities, reduced life expectancy and high prevalence of diverse, disease-associated symptoms with consequent reduced quality of life. Treatment therefor varies according to the patient's burden of disease. Aside from the presentation heterogeneity, estimating patient prognosis can be difficult. The International Prognostic Scoring System (IPSS) identifies features associated with poor outcome of PMF and can be used to estimate prognosis at the time of diagnosis, whereas the Dynamic International Prognostic Scoring System (DIPSS) can be used to inform prognosis at any time during the course of the disease. Both the IPSS and DIPSS use a series of factors that are independently associated with poor outcome, including age >65 years, haemoglobin level <100 g/l, leukocyte count >25 × 10^9/l, circulating blasts ≥1% and the presence of constitutional symptoms (fever, weight loss, and night sweats), with the DIPSS plus score incorporating thrombocytopenia, red blood cell (RBC) transfusion dependency and unfavourable karyotype as additional prognostic factors. Based on the presence of these factors, patients can be stratified according to their IPSS, DIPSS or DIPSS plus score into four risk categories: low, intermediate-1, intermediate- 2 and high-risk. Molecular prognostic markers are increasingly incorporated into risk stratification e.g. MIPSS. Based on this risk stratification, asymptomatic and low-risk patients are often observed without active treatment (the 'watch and wait' management strategy). The only curative treatment option is allogeneic haematopoietic stem cell transplantation (HSCT); however, this approach can be associated with high rates of mortality and morbidity, particularly in patients over 45 years of age and is limited to a small number of younger patients with higher risk disease. Otherise treatments are based on the goal for an individual pt: • To improve disease related systemic symptoms – ruxolitinib for INT2 & High risk pts • To improve splenomegaly and associated symptoms – ruxolitinib for INT2 & High risk pts • To improve life expectancy – ruxolitinib for INT2 & High risk pts, transplant where this is a viable option • To control myeloproliferation (raised blood counts) - hydroxycarbamide • To improve cytopenias, particularly anaemia – transfusion, erythropoietin, danazol, thalidomide

• Are any clinical guidelines used in the treatment of the condition, and if so, which?	BSH guidelines and associated update: Reilly JT, McMullin MF, Beer PA, Butt N, Conneally E, Duncombe AS, Green AR, Mikhaeel G, Gilleece MH, Knapper S, Mead AJ, Mesa RA, Sekhar M, Harrison CN. Use of JAK inhibitors in the management of myelofibrosis: a revision of the British Committee for Standards in Haematology Guidelines for Investigation and Management of Myelofibrosis 2012. Br J Haematol. 2014;167(3):418-20. Reilly JT, McMullin MF, Beer PA, Butt N, Conneally E, Duncombe A, Green AR, Michaeel NG, Gilleece MH, Hall GW, Knapper S, Mead A, Mesa RA, Sekhar M, Wilkins B, Harrison CN, Writing group: British Committee for Standards in H. Guideline for the diagnosis and management of myelofibrosis. Br J Haematol. 2012;158(4):453-71.
• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	Quite varied across UK. Ruxolitinib is becoming more established but still a lot of variation in which patients undergo active monitoring vs treatment, use of transplant, EPO, thalidomide, trials etc. We have collected (unpublished) real world data showing variation from DGH vs teaching hospital. Very much depends on the experience of the haematologist as myelofibrosis is a rare condition.
• What impact would the technology have on the current pathway of care?	I think fedratinib would have a major impact. Patients failing to respond adequately to rux (ongoing splenomegaly or disease associated symptoms) or intolerant to rux have a pretty torrid time and there is a complete lack of available treatment options currently. Availability of fed would be a major step forward
12. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Broadly yes, will be used similarly to ruxolitinib to tackle disease associated splenomegaly and symptoms in MF patients who have failed to respond adequately to rux.

• How does healthcare resource use differ between the technology and current care?	Not as much as one might think. Many patients remain on rux even if they still have symptoms or enlarged spleen as they almost always feel much worse when the treatment stops. So impact on resource might be limited as FED will replace RUX for many pts.
• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Specialist haematology clinics only.
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	None
13. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes – see above
• Do you expect the technology to increase length of life more than current care?	Probably although as pointed out, data is somewhat limited.
• Do you expect the technology to increase	Absolutely, I think this is already clear from the data in JAKARTA-2 and other parts of the submission.

health-related quality of life more than current care?	
14. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	No
The use of the technology	
15. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	Thiamine deficiency is an issue, this will require monitoring and/or routine thiamine replacement.

16. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	Very difficult to specify formal rules. Haematologists will make an evaluation per patient as to benefit versus risk of ongoing treatment
17. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	Many patients will discontinue rux in real world care so fed will be replacing rux for many pts with obvious implications to mitigate additional costs to the NHS.
18. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Not particularly innovative. It is a new class of JAK2 inhibitor
• Is the technology a 'step-change' in the management of the condition?	No, it is a new class of JAK2 inhibitor.

• Does the use of the technology address any particular unmet need of the patient population?	See above
19. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	See comment above re: thiamine deficiency. This is the main issue.
Sources of evidence	
20. Do the clinical trials on the technology reflect current UK clinical practice?	Broadly yes, I think JAKART-2 will reflect the patient group where FED is used in UK.
• If not, how could the results be extrapolated to the UK setting?	N/A
• What, in your view, are the most important outcomes, and were they measured in the trials?	Symptom and spleen responses. Survival is less clear due to study design as pointed out by the ERG.

• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	See above, spleen volume reduction is a useful surrogate of longer term improvement in survival as demonstrated for ruxolitinib in COMFORT studies.
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	Thiamine deficiency and Wernicke's encephalopathy is the main issue.
21. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No
22. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance 386?	No
23. How do data on real-world experience compare with the trial data?	No

Equality	
24a. Are there any potential equality issues that should be taken into account when considering this treatment?	No
24b. Consider whether these issues are different from issues with current care and why.	Nothing to add
Topic-specific questions	
25. For people with myelofibrosis who are relapsed/refractory to ruxolitinib, what treatments are used as part of BAT?	Detailed in above answers
26. In clinical practice, do people continue to have BAT for a lifetime? If BAT is stopped, when in the treatment pathway does this usually happen, and what treatments are offered next?	This will be assessed on a case by case basis. Typically if after 6 months there is no clear benefit then an alternative treatment should be sought, but in some cases if no alternative is available the treatment will continue. This is most important in the case of rux where patients typically continue the treatment even if response if suboptimal as they still gain significant benefit– see above.

27. Are the following groups considered to be clinically distinct populations in current NHS practice? (a) People with disease that is relapsed/refractory to ruxolitinib. (b) People who stop ruxolitinib because they cannot tolerate it or have adverse events.	Yes, these patients have a major unmet need. See responses above. Groups (a) and (b) are quite different in their profile. For example, a patient may respond well to rux but need to stop due to a side effect. Such a patient is very different (and biology of their disease also likely to be distinct) compared to a patient with ongoing splenomegaly and symptoms despite dose optimised rux.
28. In current practice, for each of the above populations, how many people would continue to have ruxolitinib, rather than stopping it and switching to something else?	80% as largely no other treatments are available outside of clinical trials. In my practice most of these patients will start experimental therapy.
29. Similarly, would you expect treatment with fedratinib to continue after relapse, rather than stopping it and switching to something else?	Yes, unless other treatment options are available.

30. If someone becomes refractory to ruxolitinib but continues to have it, would you expect them to have better outcomes compared with if they did not continue ruxolitinib?	Yes, this is very clear. Patients typically feel worse when rux stops. Of course, this does depend how one defines refractory, this is not so well established in the field and the situation will vary e.g. take two scenarios: 1. Pt with 20cm palpable spleen receives dose optimised rux for 6 months and spleen reduced to 11cm palpable and symptoms improve but are still present – is this refractory? Most criteria (including JAKARTA, FREEDOM trial criteria) would say yes, but patient will feel much worse if rux is stopped. This is the most common scenario. 2. Pt with 20cm spleen receives dose optimised rux for 6 months. Spleen increases to 26cm and patient has no improvement in symptoms. This is a much less common situation and pt is truly refractory and pt can likely stop rux without worsening of symptoms/QoL.
31. Are one or both of spleen length and spleen volume used to measure response?	Usually spleen length in routine clinical practice rather than routine MRI
32. Thinking about current NHS practice: what is the average life expectancy for people with myelofibrosis who are relapsed/refractory to ruxolitinib? What is the expected survival	In my experience survival is very poor for myelofibrosis patients who are relapsed/refractory to ruxolitinib – approx. 18-24 months unless effective experimental therapy available. Proportion alive at 5 and 10 years <5%.

proportion at 5 years and 10 years?	
33. How generalisable is the population in the JAKARTA-2 (fedratinib) trial to people with myelofibrosis who are relapsed/refractory to ruxolitinib treated in typical NHS practice? How would you expect the trial population's life expectancy to compare with that of the population you see in current NHS practice (i.e. your answer to 32)?	I think the trial criteria are reasonably generalisable by demographics and disease characteristics apart from inclusion of small number of INT-1 pts. Of course, as always, pts who are able to enter a clinical trial are a more select group.

PART 2 – Technical engagement questions for clinical experts	
Issues arising from technical engagement	
We welcome your response to the questions below, but you do not have to answer every question. If you think an issue that is important to clinicians or patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the professional organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.	
Key issue 1: Concerns with phase 2, single-arm JAKARTA-2 study	
Key issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT	

Key issue 3: Alignment between the comparator and the modelled population	
Key issue 4: Inappropriate approach to modelling	
Key issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	
Key issue 6: Inconsistent assumption between BAT and fedratinib	
Key issue 7: Assumption of survival difference	

Key issue 8: Lack of face validity for the stopping rule scenario	
Key issue 9: Costs for the comparator arm (ruxolitinib)	
Key issue 10: Reliability of response rate	
Key issue 11: End of life criteria	
Are there any important issues that have been missed in ERG report?	

PART 3 -Key messages

16. In up to 5 sentences, please summarise the key messages of your statement:

- There is a major unmet need for new therapies for myelofibrosis patients with an inadequate response to ruxolitinib
- Fedratinib has proven efficacy to improve disease associated symptoms and splenomegaly in this patient group

- Many patients would anyhow continue on ruxolitinib in this situation which will mitigate increase costs to the NHS of fedratinib being introduced for this patient group
- Additional data is needed – FREEDOM2 study is ongoing
- Thiamine deficiency is an important consideration

Thank you for your time.

Please log in to your NICE Docs account to upload your completed document, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Patient expert statement and technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

Thank you for agreeing to give us your views on this treatment and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

About this Form
In **part 1** we are asking you to complete questions about living with or caring for a patient with the condition.

In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.

The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a patient perspective could help either:

- resolve any uncertainty that has been identified
  or
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

If you have any questions or need help with completing this form please email the public involvement team via pip@nice.org.uk (please include the ID number of your appraisal in any correspondence to the PIP team).

Please return this form by **5pm** on **Friday 28 May 2021**.

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Please use this questionnaire with our hints and tips for patient experts. You can also refer to the Patient Organisation submission guide. **You do not have to answer every question** – they are prompts to guide you. There is also an opportunity to raise issues that are important to patients that you think have been missed and want to bring to the attention of the committee. The text boxes will expand as you type.

Important information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 15 pages.

PART 1 – Living with or caring for a patient with myelofibrosis and current treatment options	
About you	
1.Your name	**Mark Rutherford**
2. Are you (please tick all that apply):	☒ a patient with myelofibrosis? ☐ a patient with experience of the treatment being evaluated? ☐ a carer of a patient with myelofibrosis? ☐ a patient organisation employee or volunteer? ☐ other (please specify):
3. Name of your nominating organisation.	MPN Voice
4. Has your nominating organisation provided a submission? Please tick all options that apply.	☐ No, (please review all the questions below and provide answers where possible) ☒ Yes, my nominating organisation has provided a submission ☐ I agree with it and **do not wish to** complete a patient expert statement ☐ Yes, I authored / was a contributor to my nominating organisations submission ☐ I agree with it and **do not wish to** complete this statement ☒ I agree with it and **will be** completing

5. How did you gather the information included in your statement? (please tick all that apply)	☒ I am drawing from personal experience. ☐ I have other relevant knowledge/experience (e.g. I am drawing on others' experiences). Please specify what other experience: ☐ I have completed part 2 of the statement **after attending** the expert engagement teleconference ☐ I have completed part 2 of the statement **but was not able to attend** the expert engagement teleconference ☒ I have not completed part 2 of the statement
Living with the condition	
6. What is your experience of living with myelofibrosis? If you are a carer (for someone with myelofibrosis) please share your experience of caring for them.	I was diagnosed with polycythaemia vera in 1994. This was treated initially by venesection, then latterly with anagrelide. My polycythaemia vera progressed to myelofibrosis in 2008. Initially a watch and wait strategy was adopted. In April 2012 I started to suffer from several symptoms associated with disease progression – enlarged spleen (increased from 12 to 19cm) itchy skin, extreme fatigue, loss of appetite leading to weight loss, difficulty sleeping. I was continuing to work but from home rather than travelling. My work efficiently fell to 50-60%. Increasingly I was sleeping in the afternoon. I recognised that any further deterioration to my health would result in my giving up work. I was invited to participate in the JAKARTA trial and received my first treatment on 01.08.12. Within days my symptoms lessened – spleen reduced, appetite improved, fatigue lessened and I started to lead a normal life, working full-time. At the conclusion of the trial in November 2013 for a number of weeks I received no medication to control my myelofibrosis. During this period, the symptoms experienced prior to August 2012 returned.

Current treatment of the condition in the NHS	
7a. What do you think of the current treatments and care available for myelofibrosis on the NHS? 7b. How do your views on these current treatments compare to those of other people that you may be aware of?	I started on Ruxolitinib in late 2013 and this drug has worked well for me I'm unaware of other views on current treatments
8. If there are disadvantages for patients of **current NHS treatments** for myelofibrosis (for example how the treatment is given or taken, side effects of treatment etc) please describe these	Ruxolitinib is not effective in all patients. Additionally it can lead to weight gain. Personally I monitor closely my weight and have adjusted my diet to ensure weight gain has been minimised
Advantages of this treatment	
9a. If there are advantages of fedratinib over current treatments on the NHS please describe these. For example, the impact on your Quality of Life your ability to continue work, education, self-care, and care for others?	I believe that Fedratinib resulted in a smaller spleen than Ruxolitinib

9b. If you have stated more than one advantage, which one(s) do you consider to be the most important, and why? 9c. Does fedratinib help to overcome/address any of the listed disadvantages of current treatment that you have described in question 8? If so, please describe these.	I did not encounter excess weight gain whilst taking Fedratinib
Disadvantages of this treatment	
10. If there are disadvantages of fedratinib over current treatments on the NHS please describe these? For example, are there any risks with fedratinib? If you are concerned about any potential side affects you have heard about, please describe them and explain why.	I did not suffer any side effects from fedratinib
Patient population	
11. Are there any groups of patients who might benefit more from fedratinib or any who may benefit less? If so, please describe them and explain why.	Patients who do not respond to, or suffer significant side effects from, ruxolitinib would benefit from fedratinib. Patients who, after successful treatment, no longer respond to Ruxolitinib

Consider, for example, if patients also have other health conditions (for example difficulties with mobility, dexterity or cognitive impairments) that affect the suitability of different treatments	
Equality	
12. Are there any potential equality issues that should be taken into account when considering myelofibrosis and fedratinib? Please explain if you think any groups of people with this condition are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics More information on how NICE deals with equalities issues can be found in the NICE equality scheme	None

More general information about the Equality Act can and equalities issues can be found at https://www.gov.uk/government/publications/easy-read-the-equality-act-making-equality-real and https://www.gov.uk/discrimination-your-rights.	
Other issues	
13. Are there any other issues that you would like the committee to consider?	None

PART 2 – Technical engagement key issues for patient experts
Issues arising from technical engagement
We welcome your response to the key issues below, but you do not have to respond to every one. If you think an issue that is important to patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the patient organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.

14. **Key issue 1**: Concerns with phase 2, single-arm JAKARTA-2 study	
15. **Key issue 2**: Concerns with the unanchored indirect comparison of fedratinib to BAT	
16. **Key issue 3**: Alignment between the comparator and the modelled population	
17. **Key issue 4**: Inappropriate approach to modelling	
18. **Key issue 5**: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	

19. **Key issue 6**: Inconsistent assumption between BAT and fedratinib	
20. **Key issue 7:** Assumption of survival difference	
21. **Key issue 8:** Lack of face validity for the stopping rule scenario	
22. **Key issue 9:** Costs for the comparator arm (ruxolitinib)	
23. **Key issue 10:** Reliability of response rate	
24. **Key issue 11:** End of life criteria	
25. Are there any important issues that have been missed in ERG report?	

NICE National Institute for Health and Care Excellence

PART 3 – Key messages
26. In up to 5 sentences, please summarise the key messages of your statement: • Fedratinib alleviated all of the main symptoms I suffered from myelofibrosis • I was able to work full time and live a normal life while taking this drug • It provides a viable alternative to Ruxolitinib • There are no side effects •

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

NICE National Institute for Health and Care Excellence

Patient expert statement and technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

Thank you for agreeing to give us your views on this treatment and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

About this Form
In **part 1** we are asking you to complete questions about living with or caring for a patient with the condition.

In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.

The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a patient perspective could help either:

- resolve any uncertainty that has been identified
  or
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

If you have any questions or need help with completing this form please email the public involvement team via pip@nice.org.uk (please include the ID number of your appraisal in any correspondence to the PIP team).

Please return this form by **5pm** on **Friday 28 May 2021**.

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Please use this questionnaire with our hints and tips for patient experts. You can also refer to the Patient Organisation submission guide. **You do not have to answer every question** – they are prompts to guide you. There is also an opportunity to raise issues that are important to patients that you think have been missed and want to bring to the attention of the committee. The text boxes will expand as you type.

Important information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 15 pages.

PART 1 – Living with or caring for a patient with myelofibrosis and current treatment options	
About you	
1.Your name	Caroline Thomas
2. Are you (please tick all that apply):	☒ a patient with myelofibrosis? ☐ a patient with experience of the treatment being evaluated? ☐ a carer of a patient with myelofibrosis? ☒ a patient organisation employee or volunteer? ☐ other (please specify):
3. Name of your nominating organisation.	MPN Voice
4. Has your nominating organisation provided a submission? Please tick all options that apply.	☐ No, (please review all the questions below and provide answers where possible) ☒ Yes, my nominating organisation has provided a submission ☐ I agree with it and **do not wish to** complete a patient expert statement ☐ Yes, I authored / was a contributor to my nominating organisations submission ☐ I agree with it and **do not wish to** complete this statement ☐ I agree with it and **will be** completing

5. How did you gather the information included in your statement? (please tick all that apply)	☒ I am drawing from personal experience. ☒ I have other relevant knowledge/experience (e.g. I am drawing on others' experiences). Please specify what other experience: As a volunteer with MPS I have contact with myelofibrosis patients ☐ I have completed part 2 of the statement **after attending** the expert engagement teleconference ☐ I have completed part 2 of the statement **but was not able to attend** the expert engagement teleconference ☒ I have not completed part 2 of the statement
Living with the condition	
6. What is your experience of living with myelofibrosis? If you are a carer (for someone with myelofibrosis) please share your experience of caring for them.	I was diagnosed with Essential Thrombocythaemia (ET) in 2013. The ET has since progressed to Myelofibrosis (MF) - I have recently been diagnosed with low-risk post-ET MF, at the age of 42. I am currently, for the most part, asymptomatic, and am not taking medication. However, I have experienced intermittent symptoms, and, in my role as patient advocate, have learned a lot about the experience of living with MF, which I will describe below: MF severely reduces patients' quality of life. The cumulative burden of the large range of symptoms impacts on all aspects of our lives - family, work, social, mental and physical wellbeing, etc. This is a burden shared by all the people who make up these worlds. The most common and debilitating symptoms are as follows: <u>Fatigue</u> Most patients experience life-altering fatigue. It's important to emphasise here that fatigue doesn't mean feeling tired every now and again, or a bit run down, but instead means a patient's entire life is disrupted, and eventually governed, by a

	debilitating exhaustion that massively reduces their quality of life, often without respite. I personally have experienced fatigue so severe that I have been unable to look after my children, forcing my partner to take time off work to care for them. The stairs in my home have been insurmountable at times, meaning the really simple, unavoidable, things like changing a nappy or potty training a toddler have been a massive challenge. Perhaps the worst part is that I know I should keep myself as fit as possible, but maintaining a consistent exercise regime is impossible when the fatigue hits, leading to feelings of disappointment, guilt and generally poor mental health. The ability to exercise is important in people with cancer. Comorbidities such as obesity, high cholesterol and diabetes can result in much poorer outcomes for patients, and depression is common among cancer patients. Exercise is extremely protective against these risks, but the intense fatigue that most MF patients experience is completely incompatible with any level of physical activity. Other patients have reported a similar feeling of frustration as I have felt at being unable to maintain a healthy lifestyle due to fatigue, with the depressing addition of being trapped in a vicious circle of relying on high-fat/high-sugar foods to boost energy levels, resulting in increased risk of comorbities and thus potentially worsening the course of progression of MF. MF-induced early retirement is common amongst patients, impacting heavily on their economic status and sense of self-worth. Patients have reported having to park as close to their workplace as possible since they were unable to walk across the car park, or avoiding moving around the workplace since even low levels of activity was too tiring for them. I have heard patients describe the ‘desperate feel of heavy eyelids and limbs’ and the ‘dreadful fatigue that I can’t work with’. One patient reports that he has to lie

	down to watch TV, rather than sit, just to conserve energy. Another can no longer drive because of the fatigue. In short, fatigue in MF patients results in very poor quality of life, and the impact can contribute to poorer outcomes. <u>Enlarged spleen</u> This leads to pain, discomfort and early satiety, with accompanying unwanted weight loss. Patients' sense of well-being is often negatively impacted as they become less able to enjoy food. Body image can be negatively impacted, often resulting in distress or depression. A female patient in her 50s has described feeling 'disfigured' - her spleen protrudes through her clothes and she has lost all pride in her appearance. Male patients have talked about 'looking pregnant', and the negative psychological effects of that. Other patients report that they can no longer sleep on their front, thus affecting their sleep patterns, with further secondary impact on their lives. Patients who lead active lives are sometimes advised to give up some activities for fear of rupturing their spleen. Reduction of spleen size is a key outcome for MF patients starting a new therapy. <u>Itching</u> A very common symptom of MF is severe itching, often associated with bathing or showering. The word 'itching' doesn't really convey the extreme nature of the sensation - one patient hasn't showered or bathed for five years because the pruritus it causes is so unbearable. Another patient likens the feeling to being repeatedly rolled naked in nettles. I've heard patients describe a sensation of razor blades, mosquito bites or being poked all over with needles every second of the day.

	This extreme and constant level of irritation understandably impacts negatively on patients' relationships, work and ability to get on with normal life. One patient, a GP, says "I would frequently sit at work squirming in my chair as I tried to cope with the itching, and consult with patients at the same time". Night sweats and hot flushes Some patients have to change their bed linen during the night, on a regular basis. Other patients report frequently having to leave social engagements because the hot flushes leave them too uncomfortable or irritable. Bone pain I've heard patients compare this to a constant toothache, deep inside the bone. Painkillers are often insufficient, and the pain can severely impact on quality of life. One patient reports she can no longer drive for more than 15 minutes because of the pain in her leg bones. Poor mental health Patients have to find ways of coping with these severe symptoms, all of which can have a very negative impact on patients' mental health. The current best available therapy, Ruxolitinib, tends to reduce in efficacy after three years, and remaining on a drug that is working sub-optimally, knowing that there are no other options available, can lead to a sense of hopelessness and despair. For myself, as a mother of two young children, the knowledge that we have no second line of defence, after Ruxolitinib, is terrifying. I may be physically

	asymptomatic, but mentally, the disease takes its toll. The majority of us worry about our condition worsening. Impact on patients' communities These symptoms make life for MF patients miserable. They can also have a debilitating impact on those around them. Often, more than one member of the household has to give up their livelihood, and other members of the household may have to compensate for this. This has further knock-on effects, economically, socially and psychologically. Families also have to find a way to cope, emotionally and relationally, with one member who is in constant discomfort.
Current treatment of the condition in the NHS	
7a. What do you think of the current treatments and care available for myelofibrosis on the NHS? 7b. How do your views on these current treatments compare to those of other people that you may be aware of?	The approval of Ruxolitinib in 2015 was a big step forward - it's the only targeted therapy we have, and it's been shown to give many patients exactly what is most important to them - more life and better life. However, it's not tolerated by all patients, and it has variable and declining efficacy over time. For patients who cannot tolerate Ruxolitinib, or who have an inadequate response, there is a major unmet need. There are no approved therapies, if or when Ruxolitinib doesn't work, which reduce symptoms effectively, or extend life. Clinicians say it is better to remain on Ruxolitinib suboptimally than to try one of the previously used therapies (such as Hydroxyurea or Interferon) which are documented to be ineffective at reducing symptoms and have no survival benefit. The only cure available to us currently is a stem-cell transplant, which the majority of patients are not eligible for, and which carries high risks. In short, MF patients have one option available to us, and it's an option that we know will not work for very long.

	I'm confident that my views on current treatments are shared by the majority of MF patients.
8. If there are disadvantages for patients of **current NHS treatments** for myelofibrosis (for example how the treatment is given or taken, side effects of treatment etc) please describe these	As mentioned above, some patients (estimated at 5-10% by one clinician) have no response to Ruxolitinib. Others (estimated at a third of patients) have an inadequate response. Some cannot tolerate it at all. That's a huge proportion of our patient population who currently have no, or inadequate, treatment options. Even if Ruxolitinib provides symptom relief, it's efficacy is not long-lasting. There is no second line of defence after Ruxolitinib - our options run out and post-Ruxolitinib outcomes are dismal.
Advantages of this treatment	
9a. If there are advantages of fedratinib over current treatments on the NHS please describe these. For example, the impact on your Quality of Life your ability to continue work, education, self-care, and care for others? 9b. If you have stated more than one advantage, which one(s) do you consider to be the most important, and why? 9c. Does fedratinib help to overcome/address any of the listed disadvantages of current treatment that you	Fedratinib meets a significant unmet need for patients who have a suboptimal response to Ruxolitinib. It can be an effective option for these patients, reducing symptom burden and spleen size, thus improving quality of life. This enables patients to resume activities and continue work. Unfortunately, because trials were suspended, I have only been able to talk to one patient who has taken Fedratinib, who is also appearing as a patient expert at this appraisal. Mark will be able to provide first-hand testimony to the effectiveness of this therapy. From the perspective of an MF patient who is currently on watch and wait, one of the main advantages of Fedratinib is the psychological impact of knowing that there may be two lines of defence - it feels like a very uncertain and insecure future with only one option available. The main advantage is the reduction in spleen size and symptom burden for patients who no longer respond to Ruxolitinib.

have described in question 8? If so, please describe these.	Yes - it meets the significant unmet need of patients who do not respond, or respond inadequately, to Ruxolitinib, as well as being a second line of defence for patients whose response to Ruxolitinib begins to decline over time.
Disadvantages of this treatment	
10. If there are disadvantages of fedratinib over current treatments on the NHS please describe these? For example, are there any risks with fedratinib? If you are concerned about any potential side affects you have heard about, please describe them and explain why.	As a patient, I'm aware that no therapy is without side effects. However, Fedratinib has had positive results at trial stage which outweigh any potential disadvantages. The most serious risk can be mitigated through monitoring, and I would be prepared to accept this risk in exchange for improved quality of life and/or extended life. This is a sentiment shared by all patients I have interacted with - if the side effects are more tolerable than the symptoms of MF, which they appear to be with Fedratinib, we consider the therapy to have a favourable cost-benefit ratio.
Patient population	
11. Are there any groups of patients who might benefit more from fedratinib or any who may benefit less? If so, please describe them and explain why. Consider, for example, if patients also have other health conditions (for example difficulties with mobility, dexterity or cognitive impairments) that affect the suitability of different treatments	As mentioned, Fedratinib as a second-line treatment benefits all patients who have no or inadequate response to Ruxolitinib, or cannot tolerate it. This represents a large proportion of patients. Additionally, most patients reach a point where Ruxolitinib begins to decline in efficacy, so Fedratinib could provide a vital lifeline for these patients. I understand Fedratinib is also useful for patients with lower platelet counts. My understanding is that Fedratinib has the potential to be a useful second-line treatment for most MF patients, providing they tolerate it. It has the additional benefit of promising results as a frontline drug, showing substantial reduction in spleen size and symptom burden in a third of patients when used as a frontline therapy.

Equality	
12. Are there any potential equality issues that should be taken into account when considering myelofibrosis and fedratinib? Please explain if you think any groups of people with this condition are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics More information on how NICE deals with equalities issues can be found in the NICE equality scheme More general information about the Equality Act can and equalities issues can be found at https://www.gov.uk/government/publications/easy-read-the-equality-act-making-equality-	MF often affects older patients, which I am anxious doesn't disadvantage patients in the economic analysis.

real and https://www.gov.uk/discrimination-your-rights.	
Other issues	
13. Are there any other issues that you would like the committee to consider?	I'm very conscious that MF is a rare cancer, and that the small patient population can create uncertainty in drug trials. There is an inherent difficulty in collecting data of sufficient quantity and quality when there are so few patients to test therapies on. I'm anxious that MF patients may be disadvantaged due to our rarity, and I hope that NICE will take a flexible approach if they perceive a lack of data in support of Fedratinib. Connected to this, I'm keen that pharmaceutical companies are motivated to continue to research and develop therapies for MF patients. If approval of drugs for rare cancers is difficult to obtain purely due to the rarity, I imagine that pharmaceutical companies will invest their time and money elsewhere. Approval of Fedratinib would contribute to the evolution of targeted MPN therapy. Finally, it strikes me that medical trials, out of necessity, focus on objective, quantifiable measures. For patients, the reality is that subjective, unquantifiable measures, such as mental and physical well-being, are often more important. I have spoken to many, many patients who do not follow their 'numbers' (spleen size, blood counts) - they only really care about whether they have the energy and are comfortable enough to get out of bed in the morning and go about their day. It seems that Fedratinib could be a back-up weapon in our arsenal to allow us to do just that.

PART 2 – Technical engagement key issues for patient experts	
Issues arising from technical engagement	
We welcome your response to the key issues below, but you do not have to respond to every one. If you think an issue that is important to patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the patient organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.	
14. **Key issue 1**: Concerns with phase 2, single-arm JAKARTA-2 study	
15. **Key issue 2**: Concerns with the unanchored indirect comparison of fedratinib to BAT	
16. **Key issue 3**: Alignment between the comparator and the modelled population	

17. **Key issue 4**: Inappropriate approach to modelling	
18. **Key issue 5**: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	
19. **Key issue 6**: Inconsistent assumption between BAT and fedratinib	
20. **Key issue 7:** Assumption of survival difference	
21. **Key issue 8:** Lack of face validity for the stopping rule scenario	
22. **Key issue 9:** Costs for the comparator arm (ruxolitinib)	

23. **Key issue 10:** Reliability of response rate	
24. **Key issue 11:** End of life criteria	
25. Are there any important issues that have been missed in ERG report?	

PART 3 – Key messages

26. In up to 5 sentences, please summarise the key messages of your statement:

- The cumulative symptom burden experienced by many MF patients is massive. It is a debilitating illness that significantly reduces both length and quality of life.
- We currently only have one treatment available to us which has a significant impact on quality of life, and that impact is not felt by all patients, and in those who are lucky enough to respond well to it, it only works for a short while.
- There is a significant unmet need in a large group of patients (around 40%?) who do not respond, or who respond inadequately to Ruxolitinib.
- Fedratinib has been shown to be effective at reducing symptom burden and spleen size, and restoring quality of life to patients. It meets the unmet need of all MF patients for whom Ruxolitinib is not, or no longer, an option.

- Fedratinib would provide an important second line of defence for MF patients, whose outcomes are dismal without it. It should be approved for use.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

NICE National Institute for Health and Care Excellence

Technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments: **5pm on Thursday 5 November 2020**.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.

- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Leukaemia Care**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**n/a**

NICE National Institute for Health and Care Excellence

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: Concerns with phase 2, single-arm JAKARTA-2 study	NO	A phase 3 trial in this area is currently ongoing, FREEDOM2. Therefore, we ask that the committee consider the use of the CDF as this new trial will likely resolve these issues arising from uncertainty.
Key issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT		
Key issue 3: Alignment between the comparator and the modelled population		
Key issue 4: Inappropriate approach to modelling		
Key issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib		
Key issue 6: Inconsistent assumption between BAT and fedratinib		

Key issue 7: Assumption of survival difference	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 8: Lack of face validity for the stopping rule scenario	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 9: Costs for the comparator arm (ruxolitinib)	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 10: Reliability of response rate	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 11: End of life criteria	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses

Additional issues

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response
Additional issue 1: Unmet need	n/a	**n/a**	We would like to reiterate that this drug meets a big area of unmet need for patients.
Additional issue 2: Insert additional issue	Please indicate the section(s) of the ERG report that discuss this issue	**YES/NO**	Please include your response, including any new evidence, data or analyses, and a description of why you think this is an important issue for decision making
Additional issue N: Insert additional issue			[INSERT / DELETE ROWS AS REQUIRED]

NICE National Institute for Health and Care Excellence

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	**Company's base case before technical engagement**	**Change(s) made in response to technical engagement**	**Impact on the company's base-case ICER**
Insert key issue number and title as described in the ERG report	Briefly describe the company's original preferred assumption or analysis	Briefly describe the change(s) made in response to the ERG report	Please provide the ICER resulting from the change described (on its own), and the change from the company's original base-case ICER
..	..	..	[INSERT / DELETE ROWS AS REQUIRED]
Company's preferred base case following technical engagement	**Incremental QALYs: [QQQ]**	**Incremental costs: [£££]**	**Please provide the revised company base-case ICER resulting from combining the changes described, and the change from the company's original base-case ICER**

Technical engagement response form

Fedratinib for splenomegaly and symptoms in myelofibrosis [ID1501]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments: **5pm on Thursday 5 November 2020**.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.

- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**MPN Voice**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**none**

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: Concerns with phase 2, single-arm JAKARTA-2 study	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 3: Alignment between the comparator and the modelled population	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 4: Inappropriate approach to modelling	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 6: Inconsistent assumption between BAT and fedratinib	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses

Key issue 7: Assumption of survival difference	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 8: Lack of face validity for the stopping rule scenario	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 9: Costs for the comparator arm (ruxolitinib)	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 10: Reliability of response rate	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses
Key issue 11: End of life criteria	**YES/NO**	Please provide your response to this key issue, including any new evidence, data or analyses

Additional issues

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response
Additional issue 1: **Unmet Need**	None specific	NO	We have read the ERG report and understand, albeit from a 'informed layperson's' perspective, the issues raised. We do not have any specific responses to the issues, but we would like to take the opportunity to reiterate our central argument that Fedratinib represents a unique, effective therapy for Myelofibrosis patients for whom the effectiveness of Ruxolitinib has been exhausted. These patients have no other therapeutic options and Fedratinib therefore meets a significant unmet need for a significant proportion of Myelofibrosis patients

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	Company's base case before technical engagement	Change(s) made in response to technical engagement	Impact on the company's base-case ICER
Insert key issue number and title as described in the ERG report	Briefly describe the company's original preferred assumption or analysis	Briefly describe the change(s) made in response to the ERG report	Please provide the ICER resulting from the change described (on its own), and the change from the company's original base-case ICER
..	..	..	[INSERT / DELETE ROWS AS REQUIRED]
Company's preferred base case following technical engagement	Incremental QALYs: [QQQ]	Incremental costs: [£££]	Please provide the revised company base-case ICER resulting from combining the changes described, and the change from the company's original base-case ICER

Fedratinib for splenomegaly and symptoms in myelofibrosis. A Single Technology Appraisal

ERG commentary on the company's technical engagement response

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK Rachid Rafia, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK
Correspondence Author	Marrissa Martyn-St James, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
Date completed	27/11/2020

2. Introduction

This document sets out the ERG's commentary on the company's technical engagement response;[1] which includes a discussion of the key issues raised in the ERG report,[2] submission of new evidence and a revised economic model.

The key points raised in the company's response and the ERG's views regarding these are summarised in Table 1.

Table 1: Summary of company's technical engagement response and ERG's comments

No	Issue	Summary of company'response	ERG comment
1	Concerns with phase 2, single-arm JAKARTA-2 study	The single-arm design of JAKARTA-2 is considered justifiable for a population with a high unmet need, such as those who have failed treatment with ruxolitinib	The ERG accepts that there are differences of opinion regarding the use of single-arm studies and the lack of a concurrent control in some circumstances. However, the ERG is aware of the perspective that "the standard of care to which patients are entitled when not entered into clinical trials should be regarded as the standard by which the feasibility of the trial is judged."[3]
2	Concerns with the unanchored indirect comparison of fedratinib to BAT	Relative treatment effect estimated as odd ratios A reference is also provided to justify variable selection.	*Prognostic factors and treatment effect modifiers adjusted for in the MAICs* • The ERG does not believe that Hatswell et al (2020) addressed every scenario or the issue of variable selection. • Increasing the variance and bias might indicate that a variable is not a relevant predictor. Otherwise, the ERG suggests the joint distribution of predictor variables is what matters and that ignoring some predictor variables because they are individually balanced may not achieve balance overall.
3	Alignment between the comparator and the modelled population	The company presents new evidence on overall survival (OS) from what is refered to as an "*ongoing chart review*".	• There is a lack of information provided to conduct a full assessment of the relevance of this new source of evidence, in particular regarding the population included and patient characteristics, • The assessment relies on a naïve comparison, • It is inappropriate (biased) to compare OS from an observational study (Schain et al, 2019[4]) against OS from JAKARTA-2[5] with strict inclusion criteria (patients had to

			have a life expectancy of more than 6 month) without acknowledging differences in study populations, • SIMPLIFY-2[6] remains the most appropriate source of evidence for BAT, which is aligned with the source used for the proportion of ruxolitinib use and response rate
4	Inappropriate approach to modelling	A revised model is submitted. Structural changes are made to the model with (1) OS estimated as function of time to treatment discontinuation (TTD) and time to death following discontinution for the fedratinib arm and (2) approach to estimate TTD. Functionality included for the model to use the same random number for TTD and OS (when sampled indepently from each other).	• The model predicts implausible TTD times (some responders have a TTD time less than 24 weeks which is not possible), • The approach to parametric extrapolation of TTD that favours fedratinib is questionable, • There is a lack of detail, with some inputs not aligning with those previously reported/used in the original economic model • There are large variations in predictions for LYs between the deterministic and probabilistic analyses; highlighting the large uncertainty introduced by modelling choices made by the company, in particular with respect with the use of an individual-based approach and separating patients given to the immaturity and single-arm design of the JAKARTA-2 trial
5	Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib	Supportive care is included as a health state using data on TTD from the HMRN[7] in patients initiated on BAT and series of assumptions in patients initiated on fedratinib	• Concern with the approach taken (using OS and TTD from two different sources) • Data from the HMRN for TTD does not align with data previously reported by the company for OS in people previously treated with ruxolitinib raising questions about the source used for TTD, • A series of arbitrary assumptions are made, • It is the ERG's view that the time predicted in this health state lack plausibility

6	Inconsistent assumption between BAT and fedratinib	The company argues that there are no evidence that fedratinib would be continued following relapse	• The clinical expert interrogated as part of the technical engagement believed that following relapse, patients treated with fedratinib would remain on fedratinib, switch to ruxolitinib or move to a clinical trial.
7	Assumption of survival difference	The company comment on the appropriateness of using OS from SIMPLIFY-2[6] as OS was not published and discrepancy between the KM and the value quoted in the slide.[8] MAIC for OS between SIMPLIFY-2 and JAKARTA-2 are presented following a request from the ERG.[1]	• Exclusion of OS from SIMPLIFY-2[6] because OS has not been published and is not a valid argument; particularly in the absence of alternative evidence (Schain et al,[4] is not an appropriate source of evidence) • Studies used by the company to justify a survival difference are conducted in a different population (not receiving ruxolitinib) at the point of ruxolitinib discontinuation, rely on naïve comparisons and compare OS from an observational study against OS from JAKARTA-2[5] in which patients had to have a life expectancy of more than 6 month to be eligible for inclusion. • The comparison of OS from JAKARTA-2[5] and SIMPLIFY-2[6] up to week 24 does not support the assumption of a survival advantage for fedratinib,
8	Lack of face validity for the stopping rule scenario	The amended structure allow to better to capture the effect of the stopping rule	• There is a lack of detail about assumptions underpinning this scenario, • The ERG does not believe that a stopping rule would apply in practice because patients are left with no treatment and therefore would remain on fedratinib or switch to ruxolitinib,
9	Costs for the comparator arm (ruxolitinib)	Attempt to estimate the distribution of patients with platelet count < 100,000 x 10^9/ L	• Assuming a normal distribution leads to a large number of patients with a platelet count less than zero – a lognormal distribution is likely to be more appropriate but still with uncertainty

		based on the mean and SD in SIMPLIFY-2[6] assuming a normal distribution Reference to Gupta et al (2020)[9] regarding the dosing received in SIMPLIFY-2[6]	• It is difficult to interpret evidence from Gupta et al (2020), and the ERG notes that a proportion of patients had a mean daily dose of 0 mg, • It is challenging to estimate the cost of ruxolitinib with accuracy due its cost structure
10	Reliability of response rate	Response rate has been adjusted in an unanchored ITC	• While evidence is supportive of a better response rate with fedratinib, it is the ERG's' view that response rates cannot be considered robust as it is not possible to adjust for all differences between the SIMPLIFY-2[6] and JAKARTA-2[5] trials in terms of population, design and endpoint.
11	End of life criteria	Median LY predicted by the model in patients initiating BAT is less than 2 years	• The mean life years (LY) predicted by the model is over 2 years; with survival increasing further when more appropriate assumptions are made about the model for OS for the BAT arm.

3. ERG commentary on company's technical engagement response

Issue 1: Concerns with phase 2, single-arm JAKARTA-2 study

The ERG was concerned with the single-arm design of JAKARTA-2 as the lack of a concurrent control means that the study is likely to suffer from the phenomenon known as regression to the mean such that recruitment to the study is a consequence of extreme values that return to their average values post-treatment even if there is no treatment effect.

Issue 2: Concerns with the unanchored indirect comparison of fedratinib to BAT

The ERG does not believe that Hatswell et al (2020)[10] addressed every possible scenario or the specific issue of variable selection. Small sample sizes (and few events) makes it difficult to model even if variables are known to be prognostic. Increasing the variance and bias of an estimated treatment effect might indicate that a variable is not a relevant predictor. Otherwise, the ERG suggests the joint distribution of predictor variables is what matters, and that ignoring some predictor variables because they are individually balanced may not achieve balance overall. The ERG remains cautious about over-interpreting the results of the comparison.

Issue 3: Alignment between the comparator and the modelled population

The ERG was concerned that evidence used for the comparator arm in the economic model was taken from different sources, conducted in different populations. Indeed, in its original submission to NICE[11] the company took the response rate and the proportion of patients on ruxolitinib (89%) from the BAT arm of the SIMPLIFY-2 trial, but used overall survival (OS) from Schain et al (2019)[4] at the point of ruxolitinib discontinuation and in patients no longer treated with ruxolitinib. Therefore, OS is not aligned with the population entering the model.

In its response to the technical engagement,[1] the company included new evidence on OS from what is referred to by the company as an "*ongoing global chart review*" in patients with intermediate-2/high risk. In brief, evidence is presented for OS from the point of resistance/refractory/intolerance (r/r/i) or progression. The company states that "*The outcomes of* [redacted] *patients were reported,* [redacted] *patients had a recorded progression event and* [redacted] p*atients experienced a r/r/i event. Because the baseline for the time to OS event was taken from the point of progression if the data was available, only* [redacted] *of the patients had their baseline taken from the point of r/r/i*."

Evidence from the Chart review was included as scenario analysis only. Schain et al (2019)[4] remains the base-case source of evidence for OS for BAT.

The ERG is unclear, given the limited evidence available, why evidence from the Chart review was not presented at the time of the original submission to NICE to allow the ERG to conduct a full assessment. Limited details are provided by the company in its response to the technical engagement preventing a full assessment of the relevance of this new evidence. However, based on the limited information made available to the ERG, the ERG does not consider this new evidence to address its original concern for the following reasons:

- while it is difficult for the ERG to interpret data reported by the company because of the absence of details, the ERG does not agree with the company's statement that the study (Figure 1 in the company's technical engagement response) supports no difference in survival between patients that are continued on ruxolitinib or not. While it is difficult to interpret Figure 1 in the company's technical engagement response without any details (in particular differences in patient characteristics), the ERG notes a clear difference in survival between patients that are continued on ruxolitinib and those that did not,
- there are significant limitations (biases) with comparing survival data from an observational study (the chart review) against OS from a clinical trial with strict inclusion criteria (JAKARTA-2[5]). As highlighted in the ERG report,[2] to be recruited in JAKARTA-2, patients had to have a life expectancy of more than 6 months,

- it is unclear to the ERG what the difference is between patients classified as progressed and those that are r/r/i. The footnote underneath Figure 1 states that "*patients were considered to have progressed as defined and documented by the treating clinician*". The ERG considers this to be vague, and progression could relate to either transformation to AML, increase in spleen size or both. It is the ERG's understanding that a patient that progressed (if spleen size increase notably), would typically be considered relapsed. It is therefore unclear to the ERG whether patients with worse prognosis (compared with JAKARTA-2[5]) are included.
- No details are provided on the patients' baseline characteristics and, therefore, it is unclear whether the population included in the Chart review is similar to the one recruited in JAKARTA-2.[5] For example, it is unclear whether patient characteristics are similar in terms of age, proportion with high risk MF, proportion that are relapsed/refractory and intolerant, MF subtype, transfusion dependence. All of these characteristics would have a considerable effect on overall survival.
- OS from the Chart review is compared naively to OS from JAKARTA-2[5],

The ERG's view remains unchanged that mixing evidence from different populations and studies is not appropriate and that SIMPLIFY-2[6] is the only, and most appropriate, source of evidence for OS for the BAT arm as this is aligned with both the proportion of ruxolitinib assumed in the economic model (89%) and response rate.

Issue 4: Inappropriate approach to modelling

The ERG was concerned that the modelling approach taken by the company led to a number of biases with the ERG questioning the value of using an individual based-approach as well as separating patients onto responders and non-responders given the immaturity and single-arm nature of JAKARTA-2.[5]

In its response to the technical engagement,[1] the company agreed with the ERG that the original approach to modelling was inappropriate and led to inaccurate predictions for TTD for fedratinib. Consequently, the company submitted a revised economic model involving structural changes, with TTD now estimated from model entry, rather than using inputs split pre- and post- 24 weeks. For OS, two options have also been included that introduce dependency between OS and TTD. The first approach estimates fedratinib OS from the point of discontinuation, while the second approach uses a common random number to derive both TTD and OS events. The company also removed AML as a health state following concerns raised by the ERG.

The ERG notes that the structural changes to the model are significant and the logic of the model is different compared with the original model submitted to NICE, and therefore should be treated as a new

model.[11] Given the time and resource constraints, the ERG was unable to carry out a full assessment of the revised model submitted by the company. No details were provided about the underpinning assumptions/changes, making it challenging for the ERG to fully assess all components of the model. Nevertheless, following a brief review of the model by the ERG, while the new structure addresses some of the initial concerns (in particular around the dependency between TTD and OS), the ERG does not consider the revised structure to be more robust as a number of new issues arise. The nature and extent of these issues highlight a requirement for a full and complete assessment to ensure that the model is valid; these issues are already significant without the model being properly scrutinised:

1) The model predicts implausibe TTD for responders. TTD is estimated from model entry using a parametric model fitted to data from patients from entry into the trial. This approach led to a proportion of responder ([redacted]%; n=[redacted]) to have a TTD less than 24 weeks, despite these patients by definition being treated for at least 24 weeks. This is not possible and raises serious questions about the general model conceptualisation,
2) The ERG's understanding from the economic model (in the absence of description) is that it is assumed that [redacted]% of non-responders (including those classified as early discontinuation, early death and non-responders at 24 weeks in the original model) have a time to death equal to TTD. The company suggested that there were [redacted] TTD events for this group, with only [redacted] TTD event that was due to death. The ERG notes that this does not match data previously reported by the company. Excluding patients with clinical hold, the company previously reported that [redacted] patients with intermediate-2/high risk were available, of which [redacted] were responders at 24 weeks.[11] Therefore, the total number of non-responders (including early discontinuer [n=[redacted]], early death [n=[redacted]] and non-responders at 24 weeks [n=[redacted]]) is [redacted] patients with the number of TTD events less than this value. Indeed, data previously reported by the company as part of clarification response (KM for TTD for non-responders at 24 weeks[12]) indicated that only [redacted] patient out of [redacted] patients with no response at 24 weeks included in the KM discontinued (with the remaining [redacted] censored).
3) While minor compared with other issues, there is a strong assumption that responders at 24 weeks cannot discontinue due to death. This is because data are very immature and therefore no death was observed,
4) Despite attempts by the company to include a supportive care health state as recommended by the ERG,[2] the ERG has a number of concerns about the way this health state has been implemented as discussed in Issue 5. While the ERG understands the need for assumptions when robust data is not available, the ERG does not consider the approach taken by the company to be either appropriate nor that it generates plausible times in this health state.
5) The new approach to parametric extrapolation for TTD is highly questionable and biased in favour of fedratinib. Parametric functions are fitted to TTD data from time 0 for responders, despite these patients, by definition being alive and on treatment at 24 weeks. This approach to

parametric extrapolation will favour fedratinib. As a result, the exponential distribution is excluded. No details are provided by the company about the way survival functions were selected. In brief, the generalised gamma distribution gives the best statistical fit but is associated with a plateau. Therefore, in its base-case, the company select the lognormal distribution for TTD, the second best in terms of goodness-of-fit, for responders (fitted from Week 0), (although this is not discussed by the company). To illustrate issues with the new approach to parametric extrapolation, the ERG compared, in [redacted]1, predictions for TTD for responders used in the company base-case (lognormal distribution fitted from time zero) against predictions for TTD if parametric function were fitted from Week 24 (as patients can only discontinue after Week 24). It can be seen that some patients would discontinue before Week 24 and that this approach favours fedratinib.

[redacted]1

[redacted]

[redacted]

6) The ERG further considers the alternative approach of using the same random number to be neither robust, nor correct. Drawing from the same random number induces extreme dependency. While this makes the issue of the estimation of TTD in the original model less visible, it raises further questions about the general approach taken by the company (independent sampling) and general understanding from the company about model conceptualisation. This is not how a DES should be implemented.

Overall, the ERG's view remains unchanged that the model is overly complicated, leading to a series of conceptual errors while achieving very little benefit. It remain the ERG's view that a simpler approach (such as cohort partitioned survival model) could have been sufficient to address this decision problem given the immaturity and lack of data. While the ERG recognises that a simpler model would have some limitations and assumptions would be required (in particular around quality of life), it is the ERG'view that the value of using an individual-based model and splitting patients onto responders and non-responders leads to unnecessary complications and challenges that are difficult to reconcile.

The ERG further notes that estimates of life years (LYs) between the deterministic and probabilistic analyses in the revised model are significantly different. This is acknowledged and justified by the company because of a skewed and uncertain survival distribution notably for responders. The ERG notes that this is a consequence of modelling choices made by the company.

The company also appears to have run the PSA for 100 iterations only. For transparency, the ERG report estimates for LY for fedratinib and BAT using 1,000 iterations. The deterministic analysis predicts a mean LY of 2.84 years for BAT and 3.60 years for fedratinib vs. 3.13 years and 4.73 years in the probabilistic analysis. The ERG further notes that fedratinib is associated with less QALYs compared with BAT in 21.8% of iterations in the PSA; this is despite the model using already optimistic assumptions.

Issue 5: Omission of supportive care health state and concerns regarding HRQoL in patients initiated on fedratinib

The ERG believed that, in line with TA386,[13] supportive care should be included as a health state to reflect the period of time prior to death where the disease is no longer controlled with patients receiving only supportive treatments.

In its response to the technical engagement,[1] the company revised the model to include a supportive care health state.

While the ERG appreciates the effort and challenges involved in including this health state robustly in the absence of evidence, the ERG has a number of concerns with the approach taken by the company:

- The time in supportive care for patients initiated on BAT is estimated based on the difference between TTD and OS from two different sources. If a more appropriate approach to survival was used for BAT (same OS as for fedratinib for instance), the time in supportive care for the BAT arm would increase significantly.
- TTD from the HMRN is used.[7] The company states that "*the HMRN provided pooled second-line TTD data for patients with myelofibrosis, which was used to fit parametric extrapolations for the base case BAT TTD. A limitation of these data is that the TTD is not necessarily in relation to ruxolitinib, but to any treatment received as second-line for myelofibrosis*". The ERG is unclear whether data relate to patients previously treated with ruxolitinib only (population of interest). In particular, the ERG notes in the original CS[11] the following statement "*In the HMRN dataset, [redacted] patients discontinued treatment with ruxolitinib*". Nevertheless, the KM for the new data provided by the company appears to be based on [redacted] patients. The ERG compared OS from the HRMN in people previously treated with ruxolitinib against the TTD that is used in the revised model. While OS from HMRN is not used in the economic model, data on OS and TTD should be consistent with each other. However, it can be seen from [redacted]2 that the TTD used from the HMRN is above OS, raising significant questions about the validity of the data used by the company,
- Similarly, the time in supportive care for patients initiated on fedratinib is based on unsupported assumptions based on clinical opinion and therefore need to be considered with caution,
- Finally, it is unclear to the ERG without any details about the assumptions made for quality of life for the supportive care health state and whether it was implemented correctly. The ERG notes that the model predicted different mean utility values for patients initiated on fedratinib (0.583) or BAT (0.565) in the supportive care health state when (1) setting the discount rate for benefits to 0, (2) disabling the options to account for a decline in HrQOL with age and (3) disabling the option for worsening in quality of life.

[redacted]

The ERG further notes that data from the HMRN submitted as part of the CS indicate that following ruxolitinib discontinuation [redacted] out [redacted] patients received no further treatment, with a median survival around [redacted] months suggesting that patients would only stay in this health state for a very short time. The short time assumed in this health state was supported by the ERG's clinical experts. The company's model predicted that patients initiated on BAT and fedratinib remain in this health state for [redacted] and [redacted] years respectively raising questions about the validity of the approach taken by the company. The time would increase to about [redacted] years if patients on BAT were assumed to have the same survival as for fedratinib (ERG preferred assumption given evidence from SIMPLIFY-2[6]).

Consequently, the ERG does not consider that the updated analysis provided by the company to be appropriate or generate plausible predictions. The ERG considers that given the considerable uncertainty, it would have been more reasonable to assume that patients in both arms remain in this health state for a similar short amount of time (for instance [redacted] month), rather than estimating an arbitrary time in this health state by combining inappropriately different sources of evidence or assumptions.

Issue 6: Inconsistent assumption between BAT and fedratinib

The ERG was concerned that in its original submission to NICE,[11] the company assumed that patients continue on ruxolitinib until death (in the absence of alternative treatments), while patients initiated on fedratinib were allowed to discontinue early and move to BAT without ruxolitinib. The ERG, supported by its clinical experts, considered that should fedratinib be recommended, it is not appropriate to assume that patients who switch to fedratinib would stop treatment as no treatment is available for these patients.

In its response to the technical engagement, the company argues that the continuation of suboptimal ruxolitinib is part of the current clinical practice in the UK as opposed to suboptimal fedratinib, which is believed by the company to be an assumption in the absence of evidence of their continuation.

The ERG does not consider the argument of absence of evidence for fedratinib continuation to be valid, as fedratinib is is not currently available in the UK.

The ERG's view remains that should fedratinib be available in the UK, no alternative treatment is available for those patients following relapse and, therefore, patients are likely to continue fedratinib (as they currently do for ruxolitinib), rather than stop receiving any active treatment.

This view was also confirmed by the clinical expert contacted by NICE as part of the technical engagement process.[14] In response to what would happen to patients on fedratinib after they relapse, and whether they would continue fedratinib or stop fedratinib and switch to something else, the clinical expert stated "*I think that if patients were on fedratinib and their disease worsened or relapsed they may stay on the drug but I suspect most would switch back to Ruxolitinib or a clinical trial*."

The clinical expert for NICE's statement confirmed the ERG's view that is inappropriate to assume that following relapse, patients on fedratinib would stop fedratinib and move to BAT without ruxolitinib. Consequently, the ERG remains of the view that patients on fedratinib who relapse would either remain on fedratinib (suboptimally) or be re-treated with ruxolitinib as suggested by the clinical expert statement as part of the technical engagement.[14]

Issue 7: Assumption of survival difference

The ERG had concern that OS for fedratinib and the comparator are taken from two separate sources and that the evidence was not adjusted for differences in patients characteristics. Hence, the comparison is a naïve unanchored indirect comparison despite important differences between populations. In particular, it was the ERG's view that the studies are not directly comparable and that the population defined by Schain et al (2019)[4] does not reflect the population entering the model (consisting mostly of patients that are continued on ruxolitinib at the point of r/r/i). As such, the resulting OS predicted in the economic model does not align with evidence from SIMPLIFY-2.[6]

In its response to the technical engagement,[1] the company comments on:

(a) the appropriateness of OS from SIMPLIFY-2,[6, 8]

(b) the appropriateness of the comparison in the ERG report for OS in JAKARTA-2[5] vs. OS in Miller et al (2017)[15] and Palandri et al (2017),[16] in people with MF whom are refractory to ruxolitinib, and

(c) re-iterates arguments advanced in the CS that a survival difference is expected because of the improvement in response rate.

The company further states that the model includes the functionality to examine seven different sources of OS for BAT; all supporting a survival difference between fedratinib and BAT.

Overall, the ERG's view remains unchanged that, in light of the evidence presented by the company and identified by the ERG, no survival difference should be assumed. The model currently predicts a difference in survival at 24 weeks ([redacted]3), not backed up by evidence from SIMPLIFY-2[6, 8] (aligning with the population entering the economic model).

The ERG does not consider it a valid reason to exclude SIMPLIFY-2[6, 8] as a source of evidence for survival on the basis that evidence about OS has not been published in a peer-reviewed journal. Discrepancies between the text and the KM; as well as limitations with using data after week 24 are already acknowledged in the ERG report. It was the ERG's view that the KM should be used up to 24 weeks, rather than the value reported in the slide (as initially done in the CS).

[redacted]3[redacted]

[redacted]

[redacted]

In the ERG report,[2] a naïve comparison for OS was initially provided between JAKARTA-2[5] and SIMPLIFY-2[6, 8] at Week 24 (as patients were allowed to cross over after 24 weeks) and indicated no difference in survival before patients were allowed to cross-over.

Following the technical engagement, the ERG requested the company to provide a matched indirect comparison for OS between SIMPLIFY-2[6, 8] and JAKARTA-2,[5] to adjust for potential differences in baseline characteristics. Results were provided by the company and have been replicated below in Table 2 and

While the ERG further re-iterate that the KM from SIMPLIFY-2 up to 24 weeks should be used rather that the value in the text, the ERG notes that when adjusting for important patient characteristics ([redacted]6), OS from JAKARTA-2 at week 24 (27% probability of dying) is also below the value reported in the text (21%).

[redacted]4, [redacted]5 and [redacted]6. The analysis indicate that after matching, OS up to 24 weeks in patients initiated on fedratinib in JAKARTA-2[5] is not better than OS in a similar population treated with BAT in SIMPLIFY-2,[6, 8] and could in fact be worse when adjusting for important variables (such as platelet count and transfusion dependence).

Table 2: Exploratory OS MAICs with SIMPLIFY-2 BAT arm in first 24 weeks (reproduction of Table 3 in TE company's response)

Method	HR (95% CI)	JAKARTA-2 N / ESS

Naïve	[redacted]	[redacted]
MAIC (matching on DIPSS)	[redacted]	[redacted]
MAIC (DIPSS: matching on age and haemoglobin)	[redacted]	[redacted]
MAIC (DIPSS plus: matching on age, haemoglobin, transfusion dependence and platelet counts)	[redacted]	[redacted]

While the ERG further re-iterate that the KM from SIMPLIFY-2 up to 24 weeks should be used rather that the value in the text, the ERG notes that when adjusting for important patient characteristics ([redacted]6), OS from JAKARTA-2 at week 24 (27% probability of dying) is also below the value reported in the text (21%).

[redacted]4

[redacted])

[redacted]

[redacted]

[redacted]

In its response to the technical engagement,[1] the company also commented that the comparison provided by the ERG[2] failed to highlight that patients in JAKARTA-2[5] had a median duration of ruxolitinib exposure of [redacted] months, whilst those in Miller et al (2017)[15] and Palandri et al (2017),[16] used as supportive evidence by the ERG, had shorter duration of exposure. The ERG notes that the duration of exposure of [redacted] month referenced by the company is not specific to refractory patients, but in fact includes patients that are either relapsed or refractory (whom by definition have a longer duration of exposure to ruxolitinib).

Issue 8: Stopping rule

The ERG had concern that the original analysis done by the company only affected cost (cost were cut at week 24), but not effectiveness (OS was unchanged). The ERG also had concern regarding the validity of a stopping rule given that patients are maintained on ruxolitinib/fedratinib in the absence of alternative treatments.

In its response to the technical engagement, the company stated that the updated structure allows a better representation of the impact of the stopping rule.

No details were provided by the company regarding the different assumptions underpinning the analysis. It is therefore challenging for the ERG to provide an assessment whether the analysis is appropriate due to time and resource constraints during the technical engagement stage. Nevertheless, the ERG generally agrees with company statement that the amended structure would allow it to better capture any potential impact associated with a stopping rule.

However, the ERG re-iterates its doubt whether a stopping rule would apply in patients previously treated with ruxolitinib, as these patients are further along their treatment pathway and have no other treatment options. In particular, in relation to the company argument that patients are kept on ruxolitinib in the absence of alternative therapy. Should fedratinib be recommended by NICE, the ERG doubts that patients who switch to fedratinib be discontinued after week 24 because of the absence of response, leaving them with no other therapeutic options.

Issue 9: Cost of ruxolitinib

The ERG had concern that the proportion of patients with a baseline platelet count < 100 x 10^9/L from JAKARTA-2[5] was not representative of that in SIMPLIFY-2.[6]

In its response to the technical engagement,[1] the company attempted to estimate the proportion of patients with a platelet count below 100,000 per µl (not available in SIMPLIFY-2) based on the mean and standard deviation for platelet counts reported in SIMPLIFY-2, assuming a normal distribution. The company estimated that about 39.1% of patients in the BAT arm of SIMPLIFY-2 have a platelet count < 100 x 10^9/L. The ERG notes that this was not used in the company's updated base-case. While the ERG appreciates the effort made by the company; the ERG is concerned with assuming a normal distribution for platelet counts; which led to a large number of patients with values less than zero. The ERG notes that using a lognormal distribution (believed to be more appropriate) would increase this proportion to 49.4%. As previously highlighted in the ERG report, the proportion of patients with a platelet count < 100 x 10^9/L was 58% (median 91 x 10^9/L) in Newberry et al (2017),[17] 45% in Kuykendall, 2017,[18] and 43.5% (mean 163.9 x 10^9/L) in Palandri et al, 2019.[19]

The company further commented that "*recently published evidence from SIMPLIFY-2 from Gupta et al. has noted that 86.7% of prior JAKi-treated patients commenced ruxolitinib treatment with a mean daily dose below the maximum ruxolitinib dose, including 26.7% receiving 5 mg BID or less*". The ERG notes that the additional statement from the authors which was not reproduced in the company's response "The *mean daily dose of RUX continued to decrease, with only 5.5% of patients receiving the 20/25 mg BID maximum dose by the end of the RT period*". The ERG further notes that several patients appear to have a mean daily dose of 0 mg.

The ERG further notes that the cost structure of ruxolitinib makes it challenging to estimate a cost with accuracy. For example, it is unclear from SIMPLIFY-2 whether some people received 5 mg daily rather than 5 mg BID. In people who receive 10mg daily, it is also unclear how many patients used a single tablet (rather than two 5 mg tablets). Furthermore, it is unclear how many people received 15 mg daily (and therefore a tablet of 5mg and 10mg, or a single table of 15mg).

Overall, the ERG considers the cost for ruxolitinib to remain an area of uncertainty and one that cannot be addressed easily.

Issue 10: Reliability of response rate

The ERG's view remains unchanged about key concerns regarding the reliability of response rate due to the single arm nature of JAKARTA-2.[5] Indeed, despite attempts by the company, not all of the differences in inclusion criteria between JAKARTA-2 and SIMPLIFY-2 have been accounted for in the unanchored ITC.

It also remains the ERG's view that the difference in design between SIMPLIFY-2[6] and JAKARTA-2[5] could affect the response rate.

In JAKARTA-2,[5] all patients on fedratinib with a spleen length response also had a spleen volume reduction. The ERG notes that in SIMPLIFY-2,[6] 5.8% of patients had a spleen response based on volume, but that the response to spleen length (by palpitation) is 21%. Consequently, while the choice of response using spleen length or volume does not affect the fedratinib arm, the choice of response assessed by spleen length or volume will affect the response rate for the comparator arm.

The ERG further noted that in SIMPLIFY-2,[6] the total symptom score (TSS) was calculated based on the MPN-SAF, and this is could be less favourable to BAT as the MPN-SAF is less specific to MF symptoms.

Consequently, it is the ERG's view that while evidence is currently supportive of a better response rate for fedratinib compared with BAT, the results cannot be considered robust and that is unclear whether the difference in response rate between fedratinib and BAT currently assumed would be confirmed in FREEDOM-2 (and how many patients would receive ruxolitinib as part of BAT).

Issue 11: End of life

While the ERG recognises and re-iterates the large unmet need for this population, the ERG's view remains unchanged that it is debatable whether the population entering the economic model (as defined by the company – patients on suboptimal ruxolitinib) meets EoL criteria for the following reasons:

- the median life-years predicted by the model is less than 2 years as the company use Schain et al (2019), but this source of evidence is not considered an appropriate source by the ERG,
- the ERG further considers that the EoL criteria is not met in the company's economic model (for its base-case) even when using Schain et al (2019)[4] as it predicts a mean LY for BAT over 2 years. It should be noted that this increases even further under the ERG's preferred base-case where the same survival is assumed between treatment arms (to reflect OS from SIMPLIFY-2[6])
- It is also unclear if there is any survival gain with fedratinib given the absence of head-to-head trial data against an appropriate comparator.

4. Discussion

The ERG believes that the key issues are not resolved which relates to (a) the approach to modelling TTD as well as the general model structure/approach, (b) estimates of survival (the company approach generates a large survival gain in favour of fedratinib by week 24 contrasting with evidence from SIMPLIFY-2, and (c) the treatment pathway following relapse after fedratinib (patients either remaining on fedratinib or switching to ruxolitinib).

As illustrated in the ERG report, the ICER for fedratinib increases substantially when no survival difference is assumed (in line with evidence from SIMPLIFY-2 up to 24 weeks) and when patients initiated on fedratinib remain on treatment following relapse.

Reference

1. Celgene. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; Company's technical engagement response. 2020.
2. Martyn-St James M RR, Stevens J, Rawdin A, Wong J. Fedratinib for splenomegaly and symptoms in myelofibrosis: A Single Technology Appraisal. . *School of Health and Related Research (ScHARR)* 2020.
3. S. S. The misunderstood placebo. . *Applied Clinical Trials* 2001;10.
4. Schain F, Vago E, Song C, He J, Liwing J, Lofgren C*, et al.* Survival outcomes in myelofibrosis patients treated with ruxolitinib: A population-based cohort study in Sweden and Norway. *Eur J Haematol* 2019;103:614-9.
5. Harrison CN, Schaap N, Vannucchi AM, Kiladjian J-J, Tiu RV, Zachee P*, et al.* Janus kinase-2 inhibitor fedratinib in patients with myelofibrosis previously treated with ruxolitinib (JAKARTA-2): a single-arm, open-label, non-randomised, phase 2, multicentre study. *Lancet Haematol* 2017;4:e317-e24.
6. Harrison CN, Vannucchi AM, Platzbecker U, Cervantes F, Gupta V, Lavie D*, et al.* Momelotinib versus best available therapy in patients with myelofibrosis previously treated with ruxolitinib (SIMPLIFY 2): a randomised, open-label, phase 3 trial. *Lancet Haematol* 2018;5:e73-e81.
7. Haematological Malignancy Research N. Haematological Malignancy Research Network: Clinical Management, Resource Utilisation and Outcome in Primary and Secondary Myelofibrosis. In; 2020.
8. Oncology Sierra. Targeted Hematology and Oncology Therapeutics. 2018. http://investor.sierraoncology.com/download/Sierra+Analyst+%26+Investor+Call+Presentation_Final.pdf (Accessed 22 June 2020).
9. Gupta V EM, Hus M, et al. Momelotinib Dose-Intensity is Maintained in JAKi Naive and Previously JAKi Treated Intermediate/High-Risk Myelofibrisis Patients. *European Hematology Association* 2020;EP1103.

10. Hatswell AJ FN, Baio G. The Effects of Model Misspecification in Unanchored Matching-Adjusted Indirect Comparison: Results of a Simulation Study. *Value Health* 2020;23(6):751-759.
11. Celgene. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; 2020.
12. Celgene. Fedratinib for disease-related splenomegaly or symptoms in myelofibrosis [ID1501]. Clarification response No. 2; 2020.
13. National Institute for Health and Care Excellence. Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. Technology appraisal guidance [TA386]. 2016. https://www.nice.org.uk/guidance/TA386/chapter/1-Recommendations (Accessed 13 November 2019).
14. statement Ce. Fedratinib for splenomegaly and symptoms in myelofibrosis ID1501; Company's technical engagement response. 2020.
15. Miller CB, Komrokji RS, Mesa RA, Sun W, Montgomery M, Verstovsek S. Practical measures of clinical benefit with ruxolitinib therapy: an exploratory analysis of COMFORT-I. *Clinical Lymphoma Myeloma and Leukemia* 2017;17:479-87.
16. Palandri F, Palumbo GA, Bonifacio M, Tiribelli M, Benevolo G, Martino B*, et al.* Baseline factors associated with response to ruxolitinib: an independent study on 408 patients with myelofibrosis. *Oncotarget* 2017;8:79073-86.
17. Newberry KJ, Patel K, Masarova L, Luthra R, Manshouri T, Jabbour E*, et al.* Clonal evolution and outcomes in myelofibrosis after ruxolitinib discontinuation. *Blood* 2017;130:1125-31.
18. Kuykendall AT, Shah S, Talati C, Al Ali N, Sweet K, Padron E*, et al.* Between a rux and a hard place: evaluating salvage treatment and outcomes in myelofibrosis after ruxolitinib discontinuation. *Ann Hematol* 2018;97:435-41.
19. Palandri F, Breccia M, Bonifacio M, Polverelli N, Elli EM, Benevolo G*, et al.* Life after ruxolitinib: Reasons for discontinuation, impact of disease phase, and outcomes in 218 patients with myelofibrosis. *Cancer* 2019; 10.1002/cncr.32664.